Why Democracy Matters:
An Economic Perspective

Doctoral Thesis

to acquire the academic degree of
doctor rerum politicarum
(Doctor of Economics and Management Science)
submitted to the
School of Business and Economics of
Humboldt-Universität zu Berlin
by
MSc Vanessa Alexandra Boese

President of Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. habil. Dr. Sabine Kunst
Dean of the School of Business and Economics:
Prof. Dr. Daniel Klapper

Reviewers

1. Prof. Dr. Jochen Kluve
   (Humboldt-Universität zu Berlin, Germany)
2. Prof. Scott Gates, PhD.
   (University of Oslo, Norway)

Date of Colloquium: August 13th, 2019

Acknowledgements

Writing this thesis was a process of six years. Several people supported me in this process. Those, who provided valuable feedback for particular chapters are noted in the beginning of each chapter. In addition, there are several people whom I would like to express my gratitude toward in general:

To Jochen Kluve without whom this thesis would not have been possible.

To Scott Gates who provided invaluable guidance throughout my time as a PhD student.

To the Cusanuswerk which financed the work on my doctoral thesis through a scholarship. Thank you in particular to Liane Neubert and Dr. Martin Böke.

To my team mates at the Development Economics Network Berlin (DENeB), Katharina Lehmann-Uschner, Antonia Grohmann, Annekathrin Schoofs, Ana Garcia Hernandez and Marrit Teirlinck. Continuing the network with them has been a great experience both professionally as well as personally. I am grateful for the countless hours we spent organizing and carrying out workshops, seminars and talks while providing moral support for each other and just having a good time.

To my colleagues at the RWI Berlin office, in particular Claudia Schmiedchen, for her support and never ending supply of Agatha Christie books.

To my colleagues at the University of Oslo: Kalle Moene, Håvard Strand and Carl Henrik Knutsen, who provided some of the best comments and unfailing support. It has been both a privilege and great fun to work with you.

To Prof. Kurt Helmes who provided sound advice at critical junctures of this thesis.

To Franziska Burmeister-Naß for sharing her LaTeX-knowledge and fervor in eliminating error messages.

To Katrin Kamin. Co-authoring the two papers included in this thesis as chapters 4 and 5 made us close friends rather than co-authors, which I am very grateful for.

Last but not least, to my family (in Germany, Canada and Spain)
for their never ending love and support.

Vanessa A. Boese

Contents

0 Abstract **1**
0.1 German Abstract . . . 2

1 Introduction **3**
1.1 Why Democracy Matters . . . 3
1.1.1 The rules of the game . . . 4
1.1.2 Economic implications . . . 5
1.1.3 Structure of the thesis . . . 6

2 Viva la revolución, or: Do revolutions lead to more democracy? **7**
2.1 Introduction . . . 8
2.2 Revolutions and Democracy . . . 8
2.3 The Model . . . 9
2.4 Analysis . . . 11
2.5 Conclusion . . . 14

3 How (not) to measure democracy **15**
3.1 Introduction . . . 16
3.1.1 Definition of democracy . . . 17
3.1.2 Measurement Scale of the Democracy Measure . . . 18
3.1.3 Collection, Aggregation and Weighing of Components . . . 19
3.2 Democracy Indices . . . 21
3.2.1 Polity Index . . . 21
3.2.2 Freedom House . . . 25
3.2.3 Varieties of Democracy (V-Dem) Dataset . . . 27
3.3 Comparison . . . 31
3.3.1 Summary statistics . . . 32
3.3.2 Examining the differences . . . 33
3.4 Discussion . . . 43
3.5 Conclusion . . . 45

4 'Tis but thy name that is my enemy: on the construction of macro panel datasets in conflict and peace economics **47**
4.1 What's in a name? . . . 48
4.1.1 Typology of inconsistencies . . . 49
4.2 Democracy Data . . . 51
4.2.1 V-Dem Data Version 8 . . . 51
4.2.2 Polity IV . . . 52
4.2.3 Comparison of the democracy data . . . 53
4.3 Economic data . . . 54
4.3.1 UN Comtrade . . . 54

4.3.2 World Development Indicators . . . 55
4.3.3 Comparing the economic data . . . 55
4.4 Conflict data . . . 57
4.4.1 UCDP/PRIO Armed Conflict dataset version 18.1 . . . 58
4.4.2 MID B version 4.2 . . . 58
4.5 Discussion and conclusion . . . 59

5 Heterogeneity matters: on the dynamic interactions of trade, development, democracy and conflict 61
5.1 Introduction . . . 62
5.2 Why endogeneity matters . . . 62
5.3 Data and descriptive statistics . . . 67
5.3.1 Descriptive statistics . . . 68
5.4 Model and Estimation . . . 69
5.4.1 Reduced form VAR . . . 69
5.4.2 Structural VAR . . . 71
5.5 Results . . . 71
5.6 Robustness Checks . . . 74
5.7 Discussion & Conclusion . . . 77

Appendix 78

A Viva la revolución 81
A.1 Composition of the Pol Index . . . 81

B How (not) to measure democracy 83
B.1 Factor Analysis of the Polity2 Index components: . . . 83
B.2 Checklist of questions and respectives scores for the components of the Freedom House Index . . . 84
B.3 Setup of V-Dem's Electoral Democracy Index (Polyarchy) . . . 85
B.4 Derivation of the Polyarchy Index' rate of influence . . . 86
B.5 Summary statistics for the V-Dem democracy measures . . . 86
B.6 Democracy measures vs. "hypothetical democracy" . . . 87
B.7 Description of the democracy measures' distribution . . . 88
B.8 Examining differences in regional codings and rankings . . . 89
B.8.1 Ranking . . . 89
B.9 Eastern Europe and Central Asia (EECA) . . . 92
B.10 Countries and their respective geopolitical regions . . . 93
B.11 Slow changing level of democracy in Polity2 Index . . . 94
B.12 Countries and years included in the trunk dataset . . . 95

C 'Tis but thy name that is my enemy 99
C.1 Democracy Datasets Comparison . . . 99
C.2 Economic Datasets Comparison . . . 102
C.3 Conflict Dataset . . . 103

D Heterogeneity matters 113
D.1 List of countries and regions included in the analysis . . . 113
D.2 Regional Impulse Response Functions . . . 114
D.3 Impulse response functions - Baseline model . . . 115
D.4 Robustness Checks . . . 132
D.5 Examining residual autocorrelation . . . 139

Bibliography **140**

List of Figures

2.1 Pol Index by regions, 1960 - 2012 10

3.1 Aggregation process for high level V-Dem Indices 29
3.2 Boxplot of differences between index values for cases coded as transitions or interregnum in the Polity Scheme 34
3.3 Pairwise differences between indices plotted over average democracy levels, horizontal lines at 0, the 95th-percentile and the 5th-percentile 36
3.4 World averages of Polyarchy, Polity2 and FHI 37
3.5 Regional averages of the three democracy measures 38
3.6 Democracy coding by Polity2, FHI and Polyarchy for Albania, Bulgaria, Mongolia and Romania 39
3.7 Components of Polyarchy for Albania, Bulgaria, Mongolia and Romania 39
3.8 Democracy index values (left) and yearly rankings (right) for the United States of America from 1972-2016 41
3.9 Components of Polyarchy for the United States over time 42
3.10 Subcomponents of the clean elections index for the United States of America from 1990 - 2016 43
3.11 Average regional shares of population with suffrage, 1900 - 2016. Source: V-Dem indicator *v2x_suffr* 44

5.1 Mapping the endogenous interactions of trade, democracy, development and conflict 63
5.2 Number of countries with non/significant responses to the given shocks 73
5.3 Impulse response functions for Kenya and Portugal displaying a switch from a significant positive effect of democracy on trade to a significant negative effect and vice versa. 75

B.1 Histogram of the normalized democracy measures in trunk dataset 88
B.2 Boxplot of differences in democracy measures by geopolitical region 89
B.3 Regional average ranks based on the yearly values coded by each of the democracy indices 90
B.4 Boxplot of differences in yearly country rankings between indices by region . . . 91
B.5 Democracy coding by Polity2, FHI and Polyarchy for selected countries in Eastern Europe & Central Asia 92

D.1 Regional impulse response functions for Asia. “ac” stands for “armed conflict”, “dev” for development, and “dem” for democracy 115
D.2 Impulse response functions for countries with significant positive effects of trade on development during at least one step. 116
D.3 Impulse response functions for countries with significant negative effects of trade on development during at least one step. 117
D.4 Impulse response functions for countries with significant positive effects of development on trade during at least one step. 117

D.5 Impulse response functions for countries with significant negative effects of development on trade during at least one step. . . . 118
D.6 Impulse response functions for countries with significant positive effects of trade on democracy during at least one step. . . . 119
D.7 Impulse response functions for countries with significant negative effects of trade on democracy during at least one step. . . . 120
D.8 Impulse response functions for countries with significant positive effects of democracy on trade during at least one step. . . . 120
D.9 Impulse response functions for countries with significant negative effects of democracy on trade during at least one step. . . . 121
D.10 Impulse response functions for countries with significant positive effects of trade on conflict during at least one step. . . . 121
D.11 Impulse response functions for countries with significant negative effects of trade on conflict during at least one step. . . . 122
D.12 Impulse response functions for countries with significant positive effects of conflict on trade during at least one step. . . . 122
D.13 Impulse response functions for countries with significant negative effects of conflict on trade during at least one step. . . . 123
D.14 Impulse response functions for countries with significant positive effects of development on democracy during at least one step. . . . 124
D.15 Impulse response functions for countries with significant negative effects of development on democracy during at least one step. . . . 125
D.16 Impulse response functions for countries with significant positive effects of democracy on development during at least one step. . . . 125
D.17 Impulse response functions for countries with significant negative effects of democracy on development during at least one step. . . . 126
D.18 Impulse response functions for countries with significant positive effects of conflict on development during at least one step. . . . 127
D.19 Impulse response functions for countries with significant negative effects of conflict on development during at least one step. . . . 127
D.20 Impulse response functions for countries with significant positive effects of development on conflict during at least one step. . . . 128
D.21 Impulse response functions for countries with significant negative effects of development on conflict during at least one step. . . . 128
D.22 Impulse response functions for countries with significant positive effects of conflict on democracy during at least one step. . . . 130
D.23 Impulse response functions for countries with significant negative effects of conflict on democracy during at least one step. . . . 130
D.24 Impulse response functions for countries with significant positive effects of democracy on conflict during at least one step. . . . 131
D.25 Impulse response functions for countries with significant negative effects of democracy on conflict during at least one step. . . . 131
D.26 Robustness Check - Model in first differences: Number of countries with non/significant responses to the given shocks . . . 133
D.27 Robustness Check - Model with alternative ordering (democracy, development, trade and conflict): Number of countries with non/significant responses to the given shocks . . . 134
D.28 Robustness Check - Model with GDP per capita growth: Number of countries with non/significant responses to the given shocks . . . 135
D.29 Robustness Check - Model with logged GDP per capita: Number of countries with non/significant responses to the given shocks . . . 136

D.30 Robustness Check - Model with lag length 8: Number of countries with non/significant responses to the given shocks 137
D.31 Robustness Check - Model with lag length 10: Number of countries with non/significant responses to the given shocks 138
D.32 VAR(4): residual plots for countries with LM-test p-values between .01-.05 . . . 139
D.33 VAR(4): residual plots for countries with LM-test p-values below .01 139

List of Tables

1.1 Sample of recent studies on interrelationships of democracy and economic factors 4

2.1 Changes in system of governance . . . 9
2.2 Difference in changes in system of governance between revolution and non-revolution cases . . . 9
2.3 Results two sample t-test on changes in Pol Index . . . 9
2.4 Descriptive statistics of conflict specific variables for cases with at least one conflict per country and year . . . 11
2.5 Descriptive statistics of economic and societal variables . . . 11
2.6 Regression results using a revolution dummy, decade dummies, country fixed effects and robust standard errors (below in parentheses) . . . 12
2.7 Regression results using a duration of revolution variable, decade dummies, country fixed effects and robust standard errors (below in parentheses) . . . 13

3.1 The principles of aggregation . . . 19
3.2 Overview of the democracy measures analyzed in this paper . . . 21
3.3 Components of the Polity Index and the democratic concepts they capture . . . 22
3.4 Aggregation Rule of the Polity Index. Source: Polity IV Dataset Users' Manual, Marshall et al. (2017a), p.15, 16 and 26 . . . 23
3.5 Score to rating conversion for the Political Rights Index, Source: Freedom House (2017b) . . . 26
3.6 Pairwise correlation coefficients, bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis below. 32
3.7 Number of perfect democracies/autocracies (+/-10 in the Polity2 coding scheme) in the period 1900 - 1971 and the trunk dataset (1972 - 2015) . . . 32
3.8 Pairwise correlation coefficients for observations which are not coded "perfect" democracies/ autocracies (10/-10) in Polity2. Bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis . . . 33
3.9 Summary statistics for the democracy indices . . . 33
3.10 Results of ttest of mean differences between indices . . . 37

4.1 Number of (un)mergeable countries in a merge of the PolityIV Dataset with the COW country list . . . 53
4.2 Description of democracy datasets . . . 53
4.3 Merging V-Dem and PolityIV data . . . 54
4.4 Two sample ttests of average level of democracy. Democracy variable V-Dem: v2x_polyarchy, range: 0 (most autocratic) to 1 (most democratic) Democracy variable PolityIV: polity2, range: -10 (most autocratic) to 10 (most democratic) Note: *** Statistically significant at the 1% level. . . . 54
4.5 Description of trade datasets Note: *The total number of nonmissing observations refers to the number of observations for which the respective variable of interest contains nonmissing values. . . . 56

4.6 Merging Comtrade and WDI data Note: *When merging both datasets by country name and year those observations of inconsistency types 1-3 are unmergeable. . . 56
4.7 Two sample ttests of average level of trade and trade openness. Trade variable Comtrade: total exports, TradeValueUS, range: US$ 37,310 to US$ $2.34 \cdot 10^{16}$ Trade variable WDI: trade openness, tradeop, range: 0 to 860,8 (in %) 57
4.8 Number of (un)mergeable countries in a merge of the MID B Dataset with the COW country list 58

5.1 Selected studies on the relationships between democracy, development, trade and conflict 66
5.2 Summary statistics for pooled estimation sample 68
5.3 Variation between country means 68
5.4 Within country variation. The minimum (maximum) column refers to the lowest (highest) deviation a country exhibited from its respective panel mean. 68
5.5 Model fit for model in first differences. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported below each cell. 76
5.6 Comparing model fit across different lag specifications. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported below each cell. . 76
5.7 Comparing model fit across different specifications of development. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported in parentheses below each cell. 77

B.1 Factor analysis results for the score variables of Polity's five components 83
B.2 Checklist for the Freedom House Index. Source: Freedom House Methodology Website, Freedom House (2017b) 84
B.3 Setup of the Electoral Democracy (Polyarchy) Index. The information displayed here is gathered from the V-Dem Codebook, Coppedge et al. (2017a), p.49 - 59 and p. 435-436. 85
B.4 Summary Statistics for V-Dem democracy measures for all observations available (left) and observations in the trunk dataset (right). 86
B.5 Regression results for regression of difference between democracy measures on "hypothetical democracy". Pooled OLS, standard deviation in parenthesis below. Independent variable: "hypothetical democracy":= $\frac{(Polity2+FHI+Polyarchy)}{3}$ 87
B.6 Percentiles for Polity2 (normalized between 0-1), Polyarchy and FHI (reversed and normalized between 0-1) in trunk dataset 88
B.7 Kendall's τ_b (upper right side) and Spearman's ρ (bottom left side) 90
B.8 Countries and their respective geopolitical regions as coded in section *Comparison*. 93
B.9 Examining variation in the Polity2 Index 94
B.10 Countries (A-F) and years included in the trunk dataset 95
B.11 Countries (G-N) and years in trunk dataset continued 96
B.12 Countries (O-Z) and years in trunk dataset continued 97

C.1 Countries for which only the names/labels differ (democracy datasets) 99
C.2 Countries for which the underlying entity has no perfect match in the other dataset (Democracy Datasets) 100
C.3 Countries unmergeable due to name and time inconsistencies (democracy datasets) 101
C.4 Countries for which the names/labels differ (economic datasets) 102
C.5 Countries for which the underlying entity has no perfect match in the other dataset, A-R (economic datasets) 104

C.6 Countries for which the underlying entity has no perfect match in the other dataset, S-Z (economic datasets) 105
C.7 Countries unmergeable due to name and time inconsistencies (economic datasets) 106
C.8 Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, A-Co 107
C.9 Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, Co-Ira 108
C.10 Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, Ire-O 109
C.11 Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, P-T 110
C.12 Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, U-Z 111

D.1 Countries and regions included in the analysis 113

Chapter 0

Abstract

The current return to protectionist measures coinciding with a rise of illiberalism triggers the need for a more detailed understanding of the interactions of economic and political dimensions. This thesis consists of four articles advancing our understanding of the complex interactions between trade, democracy, development and conflict.

The first article, Boese (2015), asks: do revolutions lead to more democracy? The revolutionary conflicts examined are positively associated with a country's democratic path. In addition, the article introduces a new measure of democracy, the (X-)Pol Index.

The second article, Boese (2019), compares measure validity and reliability of Polity2, Freedom House and V-Dem democracy indices. The latter surpass the former in all relevant areas. The article provides an introduction to democracy measurement, a comparison of the advantages and disadvantages of each measure in empirical analyses and several case studies to illustrate differences across the three indices.

The third article, Boese and Kamin (2019), shows that in spite of standardization efforts the problem of inconsistent country coding across and within disciplines persists. This leads to sample selection bias as countries in conflict often undergo state name and border changes. In turn, reliability of inferences drawn from statistical analysis, in particular in conflict and peace economics, is limited. Detailed overview tables of the gravest country coding discrepancies are provided.

The fourth article (with K. Kamin, CAU Kiel) examines the interactions of democracy, development, trade and conflict. It employs a country-specific VAR to study the effects of shocks in any of the four factors on one another. Results show that these effects are vastly heterogenous across and within countries over time. The article received the Michael D. Intrilligator Best PhD Student Paper Award at the 23rd International Conference in Economics and Security in Madrid, Spain (June 2019).

0.1 German Abstract

Die derzeitige Wiederkehr protektionistischer Maßnahmen und des Illiberalismus erfordert ein detaillierteres Verständnis der Wechselwirkungen zwischen wirtschaftlichen und politischen Faktoren. Die vorliegende Doktorarbeit besteht aus vier Artikeln, die unser Verständnis der komplexen Wechselwirkungen zwischen Handel, Demokratie, Entwicklung und Konflikt voranbringen.

Der erste Artikel, Boese (2015), fragt: Führen Revolutionen zu mehr Demokratie? Die untersuchten revolutionären Konflikte sind positiv mit dem demokratischen Weg eines Landes verbunden. Darüber hinaus führt der Artikel ein neues Maß für Demokratie ein, den (X-) Pol-Index.

Der zweite Artikel, Boese (2019), vergleicht die Demokratiemaße von PolityIV, Freedom House und V-Dem. V-Dem Maße übertreffen die anderen in allen Bereichen. Der Artikel bietet eine Einführung in die Demokratiemessung, einen Vergleich der Vor- und Nachteile jedes Maßes in empirischen Analysen und Fallstudien, um die Unterschiede zwischen den Indizes zu veranschaulichen.

Der dritte Artikel, Boese and Kamin (2019), untersucht das Problem inkonsistenter Länderkodierungen zwischen verschiedenen Makrodatensätzen. Es führt zu einer Verzerrung der Stichprobenauswahl, da sich in Konfliktländern oft Name und Grenzen des Staates ändern. Dadurch wird die Zuverlässigkeit von Schlussfolgerungen aus statistischen Analysen, insbesondere in der Konfliktökonomie, eingeschränkt. Detaillierte Übersichtstabellen der Länderkodierungsdifferenzen werden bereitgestellt.

Der vierte Artikel (mit K. Kamin, CAU Kiel) untersucht die Wechselwirkungen von Demokratie, Entwicklung, Handel und Konflikt. In einem länderspezifischen VAR werden die Auswirkungen von Schocks auf einen der vier Faktoren untersucht. Die Ergebnisse zeigen, dass diese Effekte im Laufe der Zeit in und innerhalb von Ländern sehr heterogen sind. Der Artikel erhielt den Michael D. Intrilligator Best PhD Student Paper Award auf der 23. International Conference in Economics and Security in Madrid (Juni 2019).

Chapter 1

Introduction

1.1 Why Democracy Matters

A country's system of governance (or its "level of democracy"[1]) matters for economists for two reasons. The first is its relevancy due to current geopolitical developments. With the global rise of illiberalism and the recent return to protectionist measures (for example in the United States of America or China) the need for a more detailed understanding of the complex interactions of economic and political dimensions has gained renewed importance. The second reason why democracy matters is the inconclusiveness of results provided by the relevant literature. Several scholars have examined democracy and economic variables and found evidence for significant interactions between them. These results are, however, very mixed and therefore findings remain inconclusive. Table 1.1 provides a short overview of the cross-country studies most closely related to the research carried out in this thesis. In light of the first reason this raises the question why robust relationships between democracy and economic factors are so difficult to establish and what is driving the system.

My thesis contributes to an understanding of these complex interactions in the following ways: First, it shows that democracy is endogenous to economic factors such as trade openness or socio-economic development, i.e. they are jointly determined. As such it can be assumed to have direct effects on a country's economic performance as well as indirect effects through other variables (for example by decreasing the likelihood of conflict). To estimate the impact of international economic variables on each other (for example of trade openness on economic growth) it is therefore essential to account for these endogenous interactions. Second, my thesis highlights the complexity of conducting cross-national studies of observational data. Chapters 3 and 4 discuss in great detail the importance of measure validity and reliability as well as dataset construction. Third, the results of my thesis suggest that models employed in such cross-national studies need to account for economic and political factors being jointly determined and thus impacting one another both directly and indirectly. In addition, my findings suggest that these effects are very heterogenous across countries (and even within countries over time) providing a plausible explanation for the absence of robust finding in studies with models assuming homogeneity of slope parameters.

In the following I will give a brief discussion of the channels linking a country's system of governance to its economic performance. The discussion is followed by an outline of each chapter further detailing its contribution to the literature.

[1]For brevity, in this thesis the term "democracy" is used **pars pro toto**, i.e. synonymously to "system of governance" unless noted otherwise. This implies some abuse of terminology as democracy is but a certain realization on a spectrum of authority patterns. For a discussion thereof see Boese et al. (2019).

Sample studies on interactions of democracy and...	
development	
education	Murtin and Wacziarg (2014)
health (life expectancy)	Baum and Lake (2003), Besley and Kudamatsu (2006)
income inequality	Reuveny and Li (2003), Knutsen (2015)
income (GDP per capita)	Acemoglu et al. (2008), Rigobon and Rodrik (2005), Cervellati et al. (2014), Acemoglu et al. (2019), Madsen et al. (2015), Murtin and Wacziarg (2014), Narayan et al. (2011)
growth (GDP per capita growth)	Doucouliagos and Ulubaşoğlu (2008), Rodrik and Wacziarg (2005), Tavares and Wacziarg (2001), Aghion et al. (2007), Barro (1996), Knutsen (2013), Persson and Tabellini (2009), (democratisation and growth:) Papaioannou and Siourounis (2008), Giavazzi and Tabellini (2005)
economic integration	
trade openness	Li and Reuveny (2003), Reuveny and Li (2003), Rigobon and Rodrik (2005), Subramanian and Satyanath (2004), Tavares and Wacziarg (2001), Rodrik et al. (2004)

Table 1.1: Sample of recent studies on interrelationships of democracy and economic factors

1.1.1 The rules of the game

Rodrik et al. (2004) argue "what matters are the rules of the game in a society and their conduciveness to desireable economic behaviour".[2] To analyze the interactions between democracy and economic outcomes it is central to understand these rules and how they shape said interactions.

Assume a society consists of two groups, the comparatively poorer people and the wealthier elite running the country. This is but an extension of the model developed in Acemoglu and Robinson (2001) in that the poor are not necessarily disenfranchised and the elite is not necessarily autocratic. In the spirit of Baum and Lake (2003) and Lake and Baum (2001) the elite then functions as a local monopolist of public services, education etc. who "seeks to exploit their market power to produce rents that can be redistributed to either the holders of the state power",[3] i.e. the elite, or the people. In such a framework the optimal quantity supplied by an unconstrained monopolist/elite is lower than the social welfare maximizing quantity.

In this framework the central argument shaping the rules of the game in a society is that "democracy is a system in which parties lose elections" (Przeworski (1991), p.10). The degree of contestation for office, judicial and legislative constraints placed on the elite and participation of the people determines whether the monopolists produce in a contested or uncontested market, i.e. whether the elites need to fear being replaced at regular intervals. In a perfect democracy with competitive elections, full participation and constraints on the executive this will drive the monopolist elite to produce the socially optimal quantity.[4]

[2] Rodrik et al. (2004) p. 132.
[3] Baum and Lake (2003), p. 336.
[4] See Lake and Baum (2001).

1.1.2 Economic implications

This substitutability argument has far-reaching economic consequences (for example on development, growth and trade levels) but also on conflict. Broadly speaking it implies that the range of implementable policies in democratic countries is narrower that in autocratic ones. Competitive negotiation processes in democracies make resource mobilization and policy implementation more difficult. More autocratic leaders on the other hand are less bound by judicial or legislative constraints and need to fear less for their reelection (Reuveny and Li (2003)). This introduces a heterogeneity in expected policy outcomes: a strong autocrat can implement policies inducing long-term development and growth not implementable in democracies (e.g. the substantial economic growth in non-democratic China) but at the same time they can cause severe harm to the economy and the country as a whole (see for example Baschar al-Assad in Syria).

With respect to development the substitutability argument implies that democracies are more responsive to the poor who in turn push for more redistributive policies, egalitarian distributions of income and increased investments in human capital (public health and education).[5] In addition, the substitutability argument and the consequential comparative stability of policy outcomes implies indirect positive effects of democratic institutions on investments (Tavares and Wacziarg (2001)). This becomes apparent when separating the political from the economic aspects of democracy (such as protection of property rights, business and labour market regulations). These economic aspects of democratic systems are in itself growth conducive (Baum and Lake (2003)). In addition, they raise the expectations about people's returns to investments in the future exerting another positive effect on growth through investments (Persson and Tabellini (2009)).

The higher variation of policy outcomes in autocracies makes inferring a "net-effect" of a country's system of governance on its economy difficult. Even within the range of democratic countries this effect remains unclear. Tavares and Wacziarg (2001), for example, find that the increase in distributive policies in democracies comes at the expense of physical capital accumulation in turn exerting a negative effect on economic growth. While it is hard to identify such a "net-effect" the endogeneity between development and democracy is worth noting at this point. Several scholars in favour of the "modernization hypothesis" (attributed to Lipset (1959)) argue that a certain level of development acts as a prerequisite for democracy. Increased levels of socio-economic development are both outcome and determinant of democracy (Acemoglu et al. (2019), Doucouliagos and Ulubaşoğlu (2008), Madsen et al. (2015), Narayan et al. (2011)).

The "rules of the game" outlined above suggest that a country's system of governance interacts with its economy through multiple channels. Interaction here implies the possibility of causality running both ways: from democracy to the economy and vice versa. Current developments call for an increased understanding of the mechanisms how a country's economy or system of governance reacts to economic or political shocks in a globalized world. To achieve such an understanding the above mentioned dynamic interactions need to be adequately modelled and close attention needs to be paid to each of the variables included in the analysis (both in terms of theoretical justification of inclusion as much as in terms of measure validity). This is precisely what this thesis does. The following section provides a short overview of each chapter and highlights its central contribution to the literature.

[5] See Tavares and Wacziarg (2001), Doucouliagos and Ulubaşoğlu (2008), Baum and Lake (2003), Besley and Kudamatsu (2006)).

1.1.3 Structure of the thesis

The above mentioned model derived in Acemoglu and Robinson (2001) provides a possible explanation of how (threats of) revolutions, i.e. possible forced regime change, can force the elite in autocracies to democratize. This is a starting point for chapter 2.[6] It examines the effect of revolutions on a country's level of democracy. The chapter was written before the Varieties of Democracy (V-Dem) indices became available. The problems caused by a democracy measure endogenous to conflict discussed in this chapter constitute a major motivation for the research carried out in chapter 3.

The validity of any cross-country study using observational data is based upon the data. Therefore chapters 3 and 4 provide assessments of measure validity and reliability for the most frequently employed democracy measures as well as assessments of sample selection problems occurring when merging cross-country macro datasets in peace economics. It is standard to "control for" democracy in empirical cross country studies. Nevertheless, most economics studies use established yet (for certain purposes) outdated democracy indicators, such as Polity2 or Freedom House Index. These indicators are unsuitable for precisely such types of analyses on various levels (the Polity project even issues a cautionary note on the use of the Polity2 index in such studies in their Codebook, Marshall et al. (2017a) p.16). While this is a problem well know in comparative political science it has received little or no attention in economics. This is where chapter 3 comes in.[7] It compares measure validity and reliability of the established Polity2 and Freedom House indices to the comparatively new V-Dem indices. The latter surpass the former in all relevant areas. The central contribution of this chapter is its explicit and transparent comparison of the advantages and disadvantages of each measure in empirical analyses and its straightforward recommendation of using V-Dem data.

Chapter 4 discusses the implications of country naming inconsistencies between different data sets going beyond pure spelling discrepancies.[8] Properly joining economic, political and conflict data is not a trivial task as the perception of what constitutes a country/unit of analysis differs across data sets. To conduct meaningful cross-country analyses, however, the unit of analysis must be consistent across variables. The contribution of this article is twofold. First, it shows that in spite of all country coding scheme standardization efforts and relevant software packages and codes, the problem of inconsistent country coding in merging diverse macro panel datasets persists. This can lead to substantial numbers of "missing" values in merged datasets and possibly affect the reliability of inferences drawn from statistical analysis. This is of particular concern in empirical analysis in conflict and peace economics as inconsistent country coding often affects countries in conflict. Second, detailed overview tables of some of the gravest discrepancies in country coding across datasets are provided.

The last chapter constitutes a logical consequence of previous chapters. Several studies examine the interactions of democracy, socio-economic development and trade openness (see Table 1.1). Chapter 2 established an interaction between conflict and democracy. Chapter 3 and 4 provide detailed information on the choice of democracy measure and dataset aggregation. Therefore, chapter 5 estimates the interactions of democracy, development, trade openness and conflict. Its central contribution is threefold, first, the model utilized accounts for indirect effects between said factors and allows the effects to be potentially heterogeneous across countries. Second, its findings show that the four variables simultaneously affect each other and third, the observed effects are very heterogeneous across countries.

[6] Chapter 2 is published as Boese (2015).

[7] This chapter is published as Boese (2019).

[8] This chapter is joint work with Katrin Kamin, Christian-Albrechts-Universität zu Kiel and is published as Boese and Kamin (2019).

Chapter 2

Viva la revolución, or: Do revolutions lead to more democracy?

There is a vast amount of studies on the origins of revolutions and armed conflict. However, there is no empirical analysis of the political outcome of revolutionary conflicts. A second branch of research focuses on waves of democratization. This paper fills the gap and links the two fields by examining the importance of revolutions as an explanatory factor for changes in systems of governance. Using a data set covering 135 countries from 1960–2011, global changes in the level of democracy are examined. The revolutionary conflicts examined in the panel had a positive influence on a country's democratic path. In addition the effect of selected variables on the development of the system of governance is analyzed in more detail.

Keywords: conflict, democracy, democratization, polity, revolution

This chapter is published as: *Viva la Revolución, or: Do Revolutions Lead to More Democracy?* Boese, V. A. in Peace Economics, Peace Science and Public Policy, 2015: 21(4), 541-551. DOI: https://doi.org/10.1515/peps-2015-0027

Acknowledgements:

I would like to thank an anonymous reviewer, Alexander Ristig, Julien Malizard and Lloyd Jeff Dumas for excellent comments and suggestions.

2.1 Introduction

This paper attempts to answer whether and how revolutions affect countries' democratic development. Collecting macro-level data from a variety of open sources a cross-national data set covering 135 countries and the period 1960-2011 was created. While the origins of revolutions and civil conflicts have been studied extensively,[1] little theory accounting for the empirical outcome of revolutions is available. This paper uses the newly assembled data set to fill the gap by analyzing the effect of revolutions on future levels of democracy.

Building on UCDP definitions the term "revolutionary conflict" in this paper refers to *internal armed conflict occurring between the government of a state and one or more internal opposition group(s) with or without intervention from other states, in which the contested incompatibility concerns the government, i.e. the type of political system, the replacement of the central government, or the change of its composition*, see Themnér and Wallensteen (2014).

2.2 Revolutions and Democracy

This paper analyzes the development of governing structures of 135 countries over the 53 years. Hence, a "system of governance" variable was needed. A new measure, the Pol Index, was constructed based on the Polity2 Index.[2] It ranges from 0 (total loss of central political authority) to 21 (strongly democratic). Let $p_{i,t}$ denote the Pol Index of country i at time t. Then $\Delta p_{i,[t,t+k]} := p_{i,t+k} - p_{i,t}$ is the change in Pol Index for country i between the years t and $t+k$. The paper examines how a number of factors in the base year t are connected to the level of the Pol Index 2, 5 and 10 years later. It focuses on the periods $k = 2, 5, 10$ representing the short, intermediate and long run.

When dividing the dataset into two groups, observations with and without revolutionary conflict, interesting results emerge. Table 2.1 depicts the percentage of observations with improvements, no changes and deteriorations in the level of democracy 2, 5, and 10 years later for both groups. Improvements (deteriorations) refers to "substantial" positive (negative) changes in the Pol Index, i.e. changes with an absolute value larger than 3.[3] For both groups the majority of observations do not display a change in democracy level two years later. With time, the percentage of observations experiencing improvements and deteriorations sharply rises. These dynamics are even more prominent for observations with revolutions than for those without. Table 2.2 depicts the differences between the percentage values of Table 2.1 for both groups. Countries experiencing revolution had a higher volatility in their Pol Index than those without. This begs the question whether revolution cases on average might have experienced higher or lower changes in Pol Index. Using a two sample t-test it could be determined that the increase in democracy was (significantly) higher for countries undergoing revolutions in the intermediate and long run than for non-revolutionary ones, see Table 2.3.

The figures in Tables 2.1 and 2.2 raise the question whether there is a "general upwards trend" in democratization independent of revolutions. In the early 1990s Samuel Huntington observed that global patterns of democratization occur in waves.[4] This led to a wave of research on the topic.[5] Jaggers and Gurr (1995), as well as Strand et al. (2012), find evidence for a "third wave"

[1] For example by Goldstone et al. (2010) or Chenoweth and Ulfelder (2017).

[2] Detailed information on the composition of the Pol Index can be found in section A.1 of the Appendix. Polity2 Index obtained from Marshall et al. (2013).

[3] Cf. Polity IV Manual, Gurr et al. (2013), p.14.

[4] See for example Huntington (1993).

[5] See Doorenspleet (2000), Przeworski et al. (2000) or Strand et al. (2012).

				Revolution at time t			No revolution at time t		
				$k=2$	$k=5$	$k=10$	$k=2$	$k=5$	$k=10$
	$\Delta p_{i,[t,t+k]}$	>3	$\Rightarrow$ improvement	9.72%	25.77%	44.73%	3.96%	14.09%	30.29%
$-3 \leq$	$\Delta p_{i,[t,t+k]}$	≤ 3	$\Rightarrow$ no change	84.44%	64.67%	46.52%	93.67%	81.84%	64.54%
	$\Delta p_{i,[t,t+k]}$	<-3	$\Rightarrow$ deterioration	5.83%	9.56%	8.75%	2.38%	4.07%	5.17%

Table 2.1: Change in system of governance k years after time t, by percent of observations

				$k=2$	$k=5$	$k=10$
	$\Delta p_{i,[t,t+k]}$	>3	$\Rightarrow$ improvement	5.76%	11.68%	14.44%
$-3 \leq$	$\Delta p_{i,[t,t+k]}$	≤ 3	$\Rightarrow$ no change	-9.23%	-17.17%	-18.02%
	$\Delta p_{i,[t,t+k]}$	<-3	$\Rightarrow$ deterioration	3.45%	5.49%	3.58%

Table 2.2: Difference between countries with and without revolutionary conflict in base year t

Observation at time t with...	Mean change in Pol Index from t to $t+k$		
	$k=2$	$k=5$	$k=10$
...revolution	0.2934	.9596	2.3574
...no revolution	0.1651	.4527	1.0739
Difference	.1283	.5069**	1.2835***

$^{***}p<0.01, ^{**}p<0.05$

Table 2.3: Results of two sample t-test of average change in Pol Index k years later

of international democratization starting in the early 1980s.[6] In order to get an overview of trends in democratization regional averages of the Pol Index for all countries from 1960-2012 are plotted in Figure 1. Here, evidence of a decline in levels of democracy until the late 1970s and the "third wave" of democratization, is clear.

Given the positive difference in democracy changes between non-revolutionary and revolutionary observations as well as the evidence for a trend/wave of democratization, this paper contributes to the current literature by examining the importance of revolutions as an explanatory factor for changes in systems of governance.

2.3 The Model

The panel used in this paper consists of 5,786 observations from 1960-2011 and 135 countries. The explanatory variables of the model are chosen according to their established significance for democracy and political stability in recent conflict and democratization research. A binary indicator is used to mark "revolutionary" years. Countries experiencing anarchy or foreign interruption (occupation) in the base year were also flagged using dummies. Another binary variable was introduced to capture whether a country has experienced a democratic stage since 1960.[7] Additionally a *magnitude of armed conflict - score*[8] describing the destructive impact of

[6] See Jaggers and Gurr (1995), p.477.

[7] The threshold value of "democracy" here is defined as a score of at least 16 on the Pol Index.

[8] *actotal*, source: Major Episodes of Political Violence Dataset (MEPV), Marshall (2014).

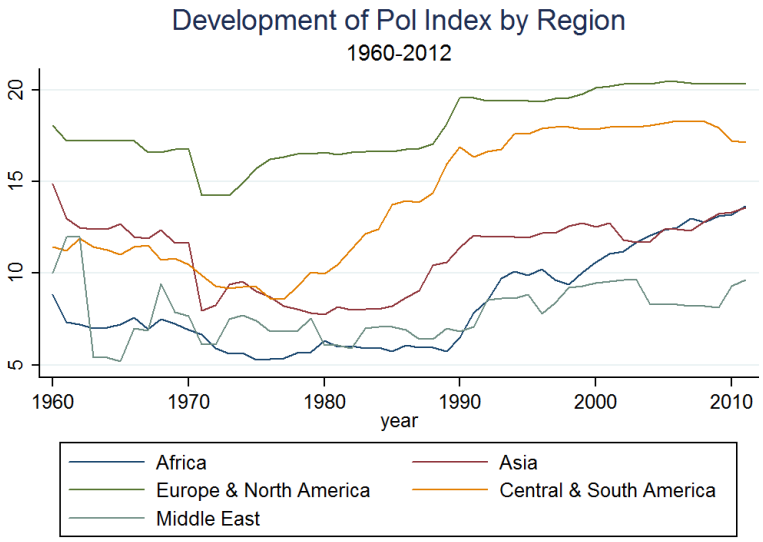


Figure 2.1: Pol Index by regions, 1960 - 2012

conflict on the country was included. In order to examine a "war weariness effect" a variable measuring the duration of armed conflict was added to the model. It counts the number of consecutive years a country has been experiencing any type of armed conflict prior to the base year. Furthermore the number of armed conflicts in the given country year was incorporated. A measure indicating the percentage of neighboring states undergoing major episodes of political violence in a given year was included.[9] To reflect the international framework, the "Affinity of Nations Index" was used.[10] Strezhnev and Voeten (2013), constructed it by using voting patterns in the United Nations General Assembly (UNGA). The voting behavior in the UNGA is regarded to be largely symbolic. Consequently this index reflects alliances/affinities between nations quite well. In particular, the index is an effective indicator of a country's alliance to superpowers. For each nation therefore its affinity to the United States, France and Russia was examined.[11] Life expectancy[12] was chosen as a proxy for the socio-economic development of a given country. From the Penn World Table, Heston et al. (2011), and the World Development Indicators, The World Bank (2014), the annual change in real GDP per capita[13] was acquired. Annual population growth (in %)[14] and primary school completion rates[15] were taken as proxies for societal conditions.

Several scholars, for example Gates et al. (2006), or Vreeland (2008), have pointed out that the participations components of the Polity2 Index can lead to an endogeneity problem when analyzing armed conflicts. This is because the participation subindices are coded with explicit reference to conflict. The Pol Index used in this paper includes the participation components as well. To show that the model's results are not driven by the endogeneity stemming from these components, the analysis was also carried out using the Pol Index without the participation

[9] Derived from NAC and NBORDER of the MEPV, Marshall (2014).
[10] *s3un*, source: Strezhnev and Voeten (2013).
[11] Due to high correlations and limited data avilability affinities to the United Kingdom and China were left out.
[12] Source: World Bank, The World Bank (2014).
[13] PPP Converted GDP Per Capita (Laspeyres), at 2005 constant prices.
[14] Source: World Bank, The World Bank (2014).
[15] *% of complete primary schooling in the age group of 15-999*, source: Barro and Lee (2013).

	Mean	Std. Dev.	Min	Max
Magnitude of destructive impact	3.491159	2.6917	0	14
% of bordering states with armed conflict	.3767415	.3219922	0	1
Conflict Duration	11.5668	12.61065	1	63

Table 2.4: Descriptive statistics of conflict specific variables for cases with at least one conflict per country and year

	Mean	Std. Dev.	Min	Max
Life expectancy, total	63.45713	11.44018	20.61488	82.93147
Population growth	1.864836	1.513716	-7.597309	17.48324
Primary school completion rates	17.8942	12.31884	.1857074	81.30415
GDP growth	2.050837	5.848191	-64.81847	90.85391
No. of armed conflicts in given country & year	.2413206	.68836	0	9

Table 2.5: Descriptive statistics of economic and societal variables, including all observations

subindices. This truncated "executive-only"-version is referred to as X-Pol Index.

Table 2.4 displays the mean, standard deviation, minimum and maximum for conflict specific variables for country-years in which at least one type of conflict occurred. Summary statistics for the socio-economic variables are presented in Table 2.5.

2.4 Analysis

The analysis was carried out using country and time fixed effects as well as robust standard errors. First, a binary revolution variable was tested to examine the effect of the occurrence of revolutions (yes or no) on future levels of democracy. The results are shown in Table 2.6. In a second step the revolution dummy was replaced by a variable counting the number of years the corresponding revolution had been ongoing. The outcome of the regression with the duration of revolution variable is shown in Table 2.7. In both tables the three columns on the right represent the analyses with the X-Pol Index.

Similarly to the findings of the t-test, in the fixed effects model revolutionary conflict seems to have a "long breath": countries with revolutions in the base year had a 0.9 higher Pol Index ten years later than "peaceful" nations. A high socio-economic development in the base year, i.e. high life expectancy and primary school completion rates, as well as low population growth, has a positive influence on the system of governance 2, 5 or even 10 years later. Interestingly, GDP growth does not have a significant effect for any number of years or in any model studied. A country's democratic lifecourse seems to be of utmost importance: those countries which had previously experienced at least one year of "democracy" had a much higher Pol Index after any number of years. An affinity to France has a very strong positive effect on the democratic development, particularly in the short run. A close affinity to Russia has a slight positive effect

Dependent Variable:	Pol Index k years later			X-Pol Index k years later		
	k=2	k=5	k=10	k=2	k=5	k=10
Revolution	-0.106	0.061	0.861***	-0.192	-0.068	0.431*
	(0.334)	(0.339)	(0.317)	(0.229)	(0.236)	(0.221)
Prior Democracy	7.850***	6.300***	4.606***	5.171***	4.153***	3.038***
	(0.256)	(0.285)	(0.284)	(0.176)	(0.195)	(0.197)
Socio-Economic Factors:						
Life Expectancy (total in years)	0.049***	0.027	-0.025	0.032***	0.016	-0.017
	(0.016)	(0.018)	(0.021)	(0.011)	(0.012)	(0.014)
Population Growth (annual %)	-0.060	-0.063	-0.191***	-0.049*	-0.052*	-0.147***
	(0.043)	(0.047)	(0.065)	(0.028)	(0.032)	(0.044)
GDP Growth	-0.008	0.002	0.014	-0.007	0.000	0.009
	(0.008)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)
% Complete Primary Schooling attained in agegroup 15-999	0.021***	0.045***	0.068***	0.010**	0.024***	0.039***
	(0.006)	(0.007)	(0.008)	(0.004)	(0.004)	(0.005)
International Factors:						
Affinity to						
USA	0.269	-0.093	-1.318***	0.085	-0.254	-0.916***
	(0.363)	(0.393)	(0.438)	(0.241)	(0.263)	(0.295)
France	1.455***	0.949*	-0.509	0.853***	0.572*	-0.421
	(0.457)	(0.496)	(0.496)	(0.306)	(0.326)	(0.340)
Russia	0.624**	0.732**	-0.012	0.473**	0.572***	0.202
	(0.285)	(0.299)	(0.325)	(0.193)	(0.202)	(0.220)
Conflict Specific Factors:						
Conflict Impact	-0.072	-0.068	-0.028	-0.052	-0.053	-0.014
	(0.050)	(0.054)	(0.056)	(0.034)	(0.037)	(0.038)
Number of Armed Conflicts	0.090	-0.142	-0.218*	0.076	-0.069	-0.106
	(0.133)	(0.125)	(0.120)	(0.088)	(0.086)	(0.084)
Conflict Duration	-0.018	-0.015	-0.006	-0.004	-0.000	0.013
	(0.014)	(0.015)	(0.016)	(0.009)	(0.010)	(0.011)
% of Bordering States with Armed Conflict	0.345	0.579**	1.647***	0.164	0.317*	1.044***
	(0.258)	(0.260)	(0.276)	(0.170)	(0.172)	(0.183)
Anarchy	-2.920***	2.014***	3.439***	-2.317***	1.173**	2.291***
	(1.014)	(0.755)	(0.755)	(0.685)	(0.543)	(0.566)
Foreign Interruption	-3.036***	0.222	0.043	-2.380***	0.054	0.445
	(0.992)	(1.025)	(0.790)	(0.685)	(0.708)	(0.533)
Decade:						
1970	-1.295***	-0.323	0.991***	-0.845***	-0.264*	0.564***
	(0.199)	(0.208)	(0.207)	(0.133)	(0.138)	(0.137)
1980	-0.100	1.557***	3.275***	-0.183	0.753***	1.818***
	(0.264)	(0.280)	(0.292)	(0.175)	(0.186)	(0.193)
1990	1.176***	2.520***	3.783***	0.513***	1.273***	2.122***
	(0.291)	(0.310)	(0.322)	(0.197)	(0.207)	(0.214)
2000	1.489***	2.714***	3.808***	0.648***	1.345***	2.110***
	(0.350)	(0.382)	(0.396)	(0.236)	(0.256)	(0.266)
2010	1.581***	-	-	0.771***	-	-
	(0.393)			(0.271)		
Number of Observations	5,876	5,480	4,815	same observations used		
% of Revolutions	10.50%	10.38%	10.34%	as on left side		
Adjusted R^2	.810	.789	.788	.798	.775	.773

$^{***}p < 0.01$,$^{**}p < 0.05$, $^{*}p < 0.10$

Table 2.6: Regression results using a revolution dummy, decade dummies, country fixed effects and robust standard errors (below in parentheses)

Dependent Variable:	Pol Index k years later			X-Pol Index k years later		
	k=2	k=5	k=10	k=2	k=5	k=10
Duration of Revolution	0.037**	0.074***	0.065***	0.019	0.046***	0.045***
	(0.017)	(0.017)	(0.019)	(0.012)	(0.012)	(0.012)
Prior Democracy	7.863***	6.313***	4.600***	5.181***	4.163***	3.036***
	(0.256)	(0.284)	(0.283)	(0.176)	(0.195)	(0.197)
Socio-Economic Factors:						
Life Expectancy (total in years)	0.050***	0.027	-0.028	0.033***	0.016	-0.019
	(0.016)	(0.018)	(0.020)	(0.011)	(0.012)	(0.014)
Population Growth (annual %)	-0.062	-0.068	-0.198***	-0.050*	-0.056*	-0.152***
	(0.043)	(0.047)	(0.065)	(0.028)	(0.032)	(0.044)
GDP Growth	-0.007	0.003	0.013	-0.006	0.001	0.009
	(0.008)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)
% Complete Primary Schooling attained in agegroup 15-999	0.021***	0.045***	0.069***	0.009**	0.024***	0.040***
	(0.006)	(0.006)	(0.008)	(0.004)	(0.004)	(0.005)
International Factors:						
Affinity to						
USA	0.286	-0.061	-1.318***	0.094	-0.233	-0.911***
	(0.363)	(0.391)	(0.435)	(0.241)	(0.262)	(0.293)
France	1.449***	0.928*	-0.548	0.857***	0.564*	-0.442
	(0.455)	(0.493)	(0.494)	(0.305)	(0.323)	(0.318)
Russia	0.619**	0.715**	-0.055	0.476**	0.566***	0.178
	(0.286)	(0.299)	(0.323)	(0.194)	(0.202)	(0.219)
Conflict Specific Factors:						
Conflict Impact	-0.076	-0.074	-0.023	-0.055	-0.058	-0.013
	(0.049)	(0.054)	(0.056)	(0.034)	(0.037)	(0.037)
Number of Armed Conflicts	0.016	-0.210*	-0.051	-0.003	-0.142*	-0.036
	(0.116)	(0.121)	(0.121)	(0.077)	(0.083)	(0.083)
Conflict Duration	-0.032**	-0.040***	-0.008	-0.012	-0.016	0.002
	(0.014)	(0.015)	(0.016)	(0.009)	(0.010)	(0.011)
% of Bordering States with Armed Conflict	0.304	0.516**	1.626***	0.136	0.273	1.021***
	(0.259)	(0.261)	(0.276)	(0.171)	(0.172)	(0.183)
Anarchy	-2.951***	1.988***	3.537***	-2.354***	1.140**	2.335***
	(1.006)	(0.749)	(0.766)	(0.680)	(0.538)	(0.570)
Foreign Interruption	-3.028***	0.358	0.387	-2.427***	0.100	0.656
	(0.983)	(1.011)	(0.808)	(0.681)	(0.697)	(0.545)
Decade:						
1970	-1.300***	-0.322	1.025***	-0.853***	-0.268*	0.581***
	(0.201)	(0.208)	(0.206)	(0.134)	(0.138)	(0.136)
1980	-0.097	1.576***	3.324***	-0.188	0.760***	1.845***
	(0.266)	(0.279)	(0.290)	(0.176)	(0.185)	(0.192)
1990	1.173***	2.539***	3.866***	0.499**	1.276***	2.166***
	(0.291)	(0.306)	(0.319)	(0.196)	(0.204)	(0.212)
2000	1.483***	2.732***	3.894***	0.631***	1.346***	2.156***
	(0.350)	(0.376)	(0.393)	(0.235)	(0.251)	(0.264)
2010	1.560***	-	-	0.743***	-	-
	(0.392)			(0.270)		
Number of Observations	5,876	5,480	4,815	same observations used		
% of Revolutions	10.50%	10.38%	10.34%	as on left side		
Adjusted R^2	.810	.790	.788	.798	.775	.773

$^{***}p < 0.01$,$^{**}p < 0.05$, $^{*}p < 0.10$

Table 2.7: Regression results using a duration of revolution variable, decade dummies, country fixed effects and robust standard errors (below in parentheses)

for the short and medium run. In this model, sharing strategic interests with the United States negatively affects the democratic development in the long run.

Two years later countries occupied by foreign powers in the base year obtained significantly lower levels of democracy than countries in charge of their own government. Those states which, during the base year, experienced a state of anarchy obtained a 2.9 lower Pol Index than those nations which did not. However, apparently "hitting rock bottom" helped in the long run: Five years later the anarchy cases had recovered and received 2.0 higher Pol Indices than the other countries. 10 years later the "anarchy" countries even had a 3.4 higher Pol Index. The findings for anarchy and foreign interruption are particularly interesting in light of the "intervention debates". For the countries studied in this model anarchy would (theoretically) be preferable to foreign interruption with respect to future democratic development. The results for the % of bordering states with armed conflict point to a "pull towards the global mean": in the long run, countries with a high percentage of neighboring states undergoing armed conflict experienced comparatively high changes towards more democracy.

Once the effect of a revolution's duration is tested the results remain similar. Experiencing an additional year of revolutionary conflict still has a positive influence on the level of democracy in the long run - but now also in the 5 year model. Taking into account the results for the X-Pol Index it becomes obvious that the models findings are not driven by the endogeneity embedded in the coding scheme of the Polity2 Index. Rather, there is stability among the results. When examining the decade dummies we find evidence for the third wave of democratization starting in the mid-1980s.

2.5 Conclusion

This paper examines the effect of revolutionary conflict on the system of governance using a panel of 135 countries from 1960-2011. The revolutionary conflicts examined in the panel had a positive influence on a country's democratic path. The results are very stable and hold even when using a truncated democracy measure evaluating only the executive dimension and thus avoiding the conflict endogeneity problem caused by the Polity IV Project's coding scheme. Additionally the model provides strong evidence for a wave of democratization starting around 1980. These findings constitute a first step in building a theory of revolution as a stimulant for democratization. Developing such a theory and including data from 2012/the Arab Spring remains an interesting task for further research.

Chapter 3

How (not) to measure democracy

Measures of democracy are regularly employed in the statistical analysis of economic, political and social policy. This paper reviews the measures' setup, strength and weaknesses across the three most prominent democracy datasets: the PolityIV, Freedom House and Varieties of Democracy (V-Dem) data. The measures developed by the V-Dem project outperform Polity2 and Freedom House Index with respect to the underlying definition, measurement scale as well as the theoretical justification of the aggregation procedure. The three indices display a high level of agreement for those observations included in all three datasets. The most substantial differences between the indices lie in the indices' coverage, i.e. in their non-missing observations (in Polity2 coding, for example, years during which a country is occupied by foreign powers constitute missing values), the availability of disaggregate data and the above mentioned key areas. This paper clarifies when to proceed with caution but for the most part advocates the use of V-Dem in the statistical analysis of democracy.

Keywords: democracy, Polity, Freedom House, Varieties of Democracy (V-Dem), measurement of democracy

This chapter is published as: *How (not) to measure democracy* Boese, V. A. in International Area Studies Review, 2019: 22(2), 95–127. DOI: https://doi.org/10.1177/2233865918815571

Acknowledgements:

I would like to thank two anonymous reviewers, Scott Gates, Carl Henrik Knutsen, Håvard Strand, Katharina Lehmann-Uschner, Steffen Elstner, Karim Diebold, and the participants of the Oxford Development Economics Workshop 2017 as well as those of the DENeB discussion round 2017 for excellent comments and suggestions.

3.1 Introduction

The characteristics of a country's system of governance are essential for the understanding of basic questions in political science and economics.[1] Over the past thirty years systems of governance have been measured in several different ways. Numerous studies have demonstrated that the results vary with the index used.[2] Hence, it is of utmost importance to understand the specific setup of the existing indices before deciding which index to use. This paper facilitates the decision which index to use when by examining the three most commonly used democracy datasets: The PolityIV, Freedom House and Varieties of Democracy (V-Dem) data. The measures developed by the V-Dem project outperform Polity2 and Freedom House Index (FHI) with respect to the underlying definition, measurement scale as well as the theoretical justification of the aggregation procedure. This article illustrates under which circumstances the three democracy measures code countries dis-/similarly and points out those frameworks under which certain measures can/should not be employed.

The question of how to correctly quantify a state's system of governance for statistical analyses has been a highly contested subject in applied political and economic research.[3] The discussion in todays literature evolves around three central questions:[4]

1. Definition: along which components should a system of governance be defined?
2. Measurement: how to quantify these components and
3. Aggregation: how to combine them into a single index of democracy?

There is no consensus in current literature on any of these questions. Due to the multi-facetted nature of the concepts "system of governance" or "democracy" a one-size-fits-all answer to the first question is unlikely to exist.[5] The characteristics of a system of governance relevant for a study vary with the research question at hand. This has led to the creation of numerous distinct democracy indices.[6] Most of the democracy indices used in the empirical literature rely on different definitions of democracy. They capture different aspects of a state's authority and decision finding structure. Several papers exist examining the strength and weaknesses of democracy measures.[7]

The suitability[8] of a democracy measure for a given statistical application depends on the research framework. However, its internal validity and reliability can be examined independently.

[1] Hence, they have been frequently examined, for example by Acemoglu et al. (2014), Franco et al. (2004), Przeworski et al. (2000).

[2] For example Bogaards (2010); Bollen (1980); Bollen and Jackman (1985), and Bollen and Jackman (1989); Casper and Tufis (2003); Cheibub et al. (2010); Elkins (2000); Treier and Jackman (2008).

[3] Boix et al. (2013), give a detailed overview of the current debates on measuring democracy, Table 1, pp.1526. Please refer to their paper for a more detailed listing of all available comments to the debate. This paper focuses on the contributions concerning Polity2, Freedom House and the V-Dem indices.

[4] Compare Munck and Verkuilen (2002), Table 2, p. 8; Boix et al. (2013), pp. 1525; Coppedge et al. (2011), p.248.

[5] Crick (2002), p.1, for example argues that the term "democracy" might be impossible to define "because the very definition carries a different social, moral or political agenda."

[6] Munck and Verkuilen (2002), Table 1, p. 6, and Table 3, p.10, present a detailed lists of available indices.

[7] For example Cheibub et al. (2010); Coppedge et al. (2008); Coppedge et al. (2011); Coppedge et al. (2017b); Elkins (2000); Gleditsch and Ward (1997); Högström (2013);Knutsen (2010); Munck and Verkuilen (2002); Munck (2009); or Treier and Jackman (2008).

[8] Suitability here refers to whether the measurement level and the democracy definition are appropriate for the given analysis.

The question "does the democracy measure capture what it is supposed to?" is generally referred to as the concept of validity. Note, that the "what it is supposed to measure"-part varies with the underlying definition of each democracy measure. "How well/ how precise does the democracy index measure what it is supposed to?" on the other hand refers to an index's reliability.[9] The two concepts are, by definition, highly interlaced. So how to assess and compare the validity and the reliability of existing democracy measures? Both, validity and reliability, depend on the underlying definition and measurement level, the data collection and aggregation procedure. Consequently, these are the central quality criterions this paper will examine for Polity2, FHI as well as for the V-Dem indices, in particular for their Electoral Democracy Index, Polyarchy. Evaluating its implicit assumptions and definitions as well as their implications for statistical analyses is crucial for choosing an appropriate democracy measure. To understand the conceptual differences underlying the definition and setup of the democracy measures each of the three central questions stated above will be briefly summarized in the remainder of this section.

3.1.1 Definition of democracy

In current debates on measures quantifying of systems of governance ("democracy indices") it has become the norm to use the term 'democracy' pars pro toto, i.e. synonymously to "system of governance". This paper follows this convention. Nevertheless, from a theoretical point of view it is important to acknowledge that the democracy indices are in fact measuring observed levels of authority patterns or systems of governance. The word 'observed' is key here: following Dahl (1971), Goertz (2006) and Treier and Jackman (2008) this paper distinguishes between the concept of democracy as a theoretical construct and its actual observable manifestations, i.e. democracy is viewed as a latent, continuous quantity. A democracy measure's validity therefore can be thought of in terms of concept-measure consistency. To obtain a valid measure of democracy, the well-defined concept should drive the methodology.[10] A consequence of this latency is the uncertainty/noise in coding of the empirical measures.[11] In addition to noise in the measures coding due to the latency of democracy several studies, such as Bollen (1993), Bollen and Paxton (2000) or Elff and Ziaja (2018), show that method factors (i.e. coder-specific systematic biases in each dataset) can provide further sources of uncertainty in democracy measures. To avoid false inference a precise definition of the concept of democracy[12] guiding the methodological construction of the empirical index is essential. Following Goertz (2006) and Dahl (1971) perfect democracy and autocracy can be thought of as the poles between which a continuum of polyarchies/authority patterns/systems of governance unfolds.

Even on a conceptual level the question of how to conceptualize a system of governance is an entire field of research.[13] Schumpeter's minimalist definition[14] captures the core trait of an institutional definition of democracy: **a regime holding contested elections - contestation and participation**.[15] It is worth noting, that these two democratic principles are complements: a country cannot be considered democratic if there is full participation, but no contestation and

[9] For a discussion of validity and reliability see, for example, Coppedge et al. (2017b), pp.16; and Munck (2009), pp. 23.

[10] See Goertz (2006), p.11.

[11] Treier and Jackman (2008) show this for the Polity index and Høyland et al. (2012) show it for index rankings using (among others) the FHI.

[12] The point cannot be stressed enough that this precise definition of the concept might vary with application.

[13] See for example Schmitter and Karl (1991); Crick (2002); Huntington (1993), pp.5; Diamond (1999), pp. 7; or Dahl (1971), chapter 1.

[14] "(...) the democratic method is that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote" Schumpeter (1976 [1942]), p.269.

[15] Participation here is used in an inclusiveness/suffrage sense.

vice versa.

The central criticism of this minimalist institutional definition is that by acknowledging only the electoral dimension other defining components of democracy are ignored.[16] Instead of defining democracy institutionally Dahl (1971) goes back to the word's greek origin.[17] He takes a substantive approach and views **popular control over collective decision making as well as political equality** as core elements of democracy.[18] He recognizes that to realize this democratic principle of popular rule and political equality a number of institutional guarantees have to be met. These institutional guarantees are a necessary (not sufficient!) condition to ensure the implementation of the two core institutional democracy traits of contestation and participation. In that sense there cannot be a democratic regime without holding contested elections but at the same time not all regimes holding contested elections classify as democracies.[19]

On a conceptual level, a distinction between *de facto* and *de jure* attributes of a polity can be made. This paper regards the formal institutional setup of a polity as its *de jure* framework. Observable policy outcomes on the other hand constitute *de facto* elements. This concept of *de facto* encompasses the way the institutions operate in practice as well as the performance and practices generated by the actors in the system.[20] Political violence carried out by non-governmental actors for example can be an byproduct of the "institutional output" unintended by the system's setup and consequently be part of the *de facto* reality.

Lastly, it is not enough to identify components to include in a definition of democracy. The theoretical relations of these components with each other need to be taken into account. Only once a logical structure between these elements is established they can be aggregated into a measure of democracy in a meaningful way.[21] Not ordering the elements vertically by their level of abstraction gives rise to problems of redundancy and conflation (joining elements which are symptoms of distinct overarching attributes).[22]

3.1.2 Measurement Scale of the Democracy Measure

The question how to quantify the components and on what kind of scale to map the empirical distribution of cases should also be determined by the concept behind the measure. Most polychotomous democracy indices, such as Polity2 or FHI, are coded on a nominal or ordinal scale. However, in empirical research these indices are often treated as if they were coded on an interval scale even though there is no numerical relationship between the categories. Doing so implies that the differences between the values have a meaningful interpretation. This is a major point of criticism of several of the polychotomous indices.[23] While the definition and aggregation procedure of the V-Dem democracy indices[24] permits using them as quasi continuous measures,

[16] See, for example, Diamond (1999), p.9; Mainwaring et al. (2007), p.128; Schmitter and Karl (1991), p. 78.

[17] "$\delta\eta\mu o\kappa\rho\alpha\tau\acute{\iota}\alpha$" is a compound of $\delta\tilde{\eta}\mu o\sigma$ - the comon people, citizens and $\kappa\rho\acute{\alpha}\tau o\sigma$ - rule; thus literally meaning "rule of the people".

[18] See Dahl (1971), p.1.

[19] See Cheibub et al. (2010), p. 72.

[20] An anonymous referee's comments were very helpful in substantiating this definition and are very much appreciated.

[21] Munck and Verkuilen (2002) p. 13, provide an overview of such a logical organization of elements.

[22] Munck and Verkuilen (2002), for example, point out that the Polity IV Data's aggregation rule suffers from a redundancy problem, p.13.

[23] See for example Gleditsch and Ward (1997), p. 380, point 1; Cheibub et al. (2010), p. 75 (FHI), p. 76 (Polity Index); Boix et al. (2013), p. 1529.

[24] They are aggregated using a mixture of Bayesian item response theory measurement models, addition and multiplication (see section *Varieties of Democracy (V-Dem) Dataset*).

1. Data Collection Process: type of information used
    - clear and transparent rules for the collection of information
    - public availability of disaggregate data
    - sources: subjective evaluations vs. objective facts
2. Aggregation Rule: how information is aggregated and weighted
    - choice of level of aggregation
    - capture of underlying theory (reflecting an accepted definition of democracy) in the aggregation rule
    - Provision and justification of the aggregation rule
        - substitutability vs. complementary relations
        - factoranalysis or principal component analysis
        - addition, multiplication
        - openness to further tests

Table 3.1: The principles of aggregation

this is not the case for Polity2 and FHI. Polity2 is defined on a scale which is not even ordinal,[25] and differences between categories have no meaningful interpretation. One specific FHI score can be the result of a variety of different underlying factors. The Freedom House aggregation procedure using equal weighing and collapsing further contributes to its final measurement scale at best being ordinal.

Concept-measure consistency is also essential in terms of the scale's end points. Whether a democracy index can take on top- or bottom values should be determined by whether the concept of democracy allows for perfect democracy and autocracy endpoints. Assuming a continuous concept the ends of the conceptual scale can be thought of as unreachable poles (one can get infinitely close to but never reach them). Fixed endpoints imply that one believes in a state of perfect autocracy/democracy. From the empirical point of view there is a disadvantage to attainable top- and bottom values which becomes obvious with long time series. For example, the increased number of democracies in the system of states after the end of the Cold War made it impossible to examine "the societal dynamics associated with the consolidation and maintenance of democratic authority patterns"[26] within these states using the Polity Coding Scheme.

3.1.3 Collection, Aggregation and Weighing of Components

The question of how to adequately collect, weigh and aggregate components has been subject to much debate.[27] The most important discussion points are displayed in Figure 3.1. The validity and reliability of a measure increase the more adequately these points are adressed.

[25] See section *Polity Index*. Cases of interregnum/anarchy, for example, are coded as 0 - the "middle" of perfect democracy and perfect autocracy.

[26] See Marshall et al. (2017a), p.10.

[27] See for example Cheibub et al. (2010), pp 74;Coppedge et al. (2011), pp. 250; Boix et al. (2013) p. 1530; Knutsen (2011), pp. 83; Munck and Verkuilen (2002), pp. 22.

The first principle of aggregation acknowledges the fact that the type of information captured in the disaggregated data is essential.[28] Several democracy measures are based on components coded using subjective inferences "and perhaps even guesses", Cheibub et al. (2010), p. 77. Including of subjective evaluations reduces replicability and increases measurement error, but due to democracy being a social construct a limitation to formal (and better quantifyable) indicators leads to lower validity.[29] Replicability is a key feature for a useful democracy measure. Hence, clear and transparent rules for the information collection process and public availability of disaggregate data should be considered crucial.[30] This point has not been adequately addressed by many of the existing indices, e.g. the FHI, and consequently has been frequently criticised.[31]

The second principle of aggregation concerns the Aggregation Rule, i.e. how the disaggregate information is weighted and aggregated. First, a suitable level of aggregation should be determined.[32] While a high level of aggregation (aggregating all components and indicators into one democracy index) is appealing for the use of a democracy index in empirical calculations it also entails loss of complexity, information and validity.[33] Second, the underlying theory (reflecting an accepted definition of democracy)[34] should be captured in the aggregation rule.[35] Third, to increase replicability the aggregation rule should be provided and justified.[36] This entails for example a justification of choice of relation between the components: are they substitutes or complements? A non-arbritrary aggregation rule justifies its use of addition or multiplication of components or even the use of factor or principal component analyses. And, last but not least, it is open to further examinations.[37]

Democracy as a political concept is inherently difficult to define and to capture in a quantitative way. Yet, this paper demonstrates that the V-Dem project succeeded in constructing quantitative measures excelling in the key areas mentioned above. The goal of this paper is thus to provide a thorough overview of relevant issues to consider when choosing a democracy measure for a statistical application. It seeks to give the reader an understanding of the differences, weaknesses and strengths of the three democracy datasets.

The following three points distinguish this article from the existing literature comparing democracy measures and their performance:

1. It is up to date: includes a discussion of the V-Dem indices which only became available in 2014

2. It shows that the biggest differences between the democracy measures lie in their coverage, their underlying definition, measurement scale, the justification of their aggregation procedure as well as the availability of disaggregate data and explains under which conditions certain measures can/should not be used.

3. Gathering this information in one article reduces the search cost for scholars working with democracy measures.

[28] Cf. Cheibub et al. (2010), p. 74.
[29] Cf. Knutsen (2011), p. 84.
[30] Cf. Cheibub et al. (2010), p. 74; Munck and Verkuilen (2002), Munck and Verkuilen (2002), p. 25 & 26.
[31] For example by Cheibub et al. (2010), pp. 75, or Munck and Verkuilen (2002), Munck and Verkuilen (2002), p. 25.
[32] Cf. Munck and Verkuilen (2002), p. 23.
[33] Cf. Munck and Verkuilen (2002), p. 22.
[34] Cf. Coppedge et al. (2011), p. 250.
[35] Cf. Knutsen (2011) p. 83.
[36] See also Munck and Verkuilen (2002).
[37] Cf. Munck and Verkuilen (2002), p. 25.

Index	Definition includes	Availability from	to	N	Range	Aggregation rule	Advantages	Disadvantages
Freedom House Index	Subjective lists of civil liberties and political rights	1972	2016	209	1 (highest) to 7 (lowest)	Addition, ranking (including collapsing variation) and averaging	- Covers a variety of democratic features (civil liberties and political rights)	- Maximalist (subjective) democracy definition - Definition changed over time - Aggregation rule without justification - Measurement problems - Disaggregate data not publicly available - Unclear distinction between de facto and de jure aspects of a polity and their influence in the democracy measures
Polity2 Index	Contestation of offices, constitutional constraints on head of state	1800	2016	195	-10 (lowest) to 10 (highest)	Combination of weighing and addition	- Broad temporal & spatial scope - Detailed coding rules - Disaggregate data publicly available	- No theory behind aggregation rule - Definition: omission of suffrage/any participation - Factionalism categorization - 0 coding for interregnum - Missing values for foreign interruption
V-Dem Democracy Indices	Electoral, liberal, egalitarian, deliberative, and participatory dimension	1900	2016	177	0 (lowest) to 1 (highest)	Mixture of Bayesian item response theory measurement models, addition and multiplication	- Broad temporal & spatial scope - Weakest link argument included in aggregation procedure - Theoretical justification of aggregation rules - Bridge- and Lateral-coding - Disaggregate data publicly available - Public discussion of measurement error - Ordinal versions of all variables offered in addition	- Unclear distinction between de facto and de jure aspects of a polity and their influence in the democracy measures

N = number of countries/territories coded. Range: (lowest/highest) refers to the lowest/highest possible level of democracy

Table 3.2: Overview of the democracy measures analyzed in this paper

3.2 Democracy Indices

The number of existing democracy indices is too vast to give a detailed overview in one paper.[38] Some indices, such as the V-Dem indices are available in country-event format. For the sake of increased comparability among indices and the usability in classic time-series cross-country studies this paper focuses on data in country-year format. Table 3.2 summarizes the three main democracy measures discussed in this section. It displays the measurement scale, the democracy definition, the measure's temporal and geographical scope, its' range, aggregation rule as well as its' strengths and weaknesses.

3.2.1 Polity Index

Due to its broad chronological (1800 - 2016) and geographical scope (195 countries) the Polity Index is one of the most frequently used democracy indices in current research. When it was first introduced in 1975 it constituted one of the first major attempts to quantify authority patterns on a global scale over an extended period of time. These first codings were based on the detailed theories of authority patterns put forward by Eckstein (1973) and Gurr (1974). The first Polity data was coded with the explicit objective of answering the question whether "the durability (persistence and adaptability) of political systems depends at all upon the nature of their structures of political authority".[39] Said structures of political authority, i.e. the institutional framework therefore form the core of the Polity Index.

[38] For a quick overview please refer to Coppedge et al. (2017b), Table 1; Munck (2009), Table 4.2 or Pemstein et al. (2010), Table 1.

[39] See Gurr (1974), p.1482.

Competitiveness of Executive Recruitment	XRCOMP	contestation
Openness of Executive Recruitment	XROPEN	contestation
Constraint on Chief Executive	XCONST	institutional constraints
Competitiveness of Political Participation	PARCOMP	contestation
Regulation of participation	PARREG	contestation

Table 3.3: Components of the Polity Index and the democratic concepts they capture

Today the data is assembled by researchers from the Polity IV Project[40] at the Center for Systemic Peace. The most recent version of data available is the Polity IV Dataset Version 2016, Marshall et al. (2017b).[41]

The Polity Index ranges from -10 (strongly autocratic) to 10 (strongly democratic) and is calculated by subtracting a measure of autocracy (Autoc) from a measure of institutionalized democracy (Democ): $Polity = Democ - Autoc$. Both indices, Democ and Autoc, range from 0 to 10. They are made up of scores of five different components reflecting the polity's executive recruitment (XRCOMP[42] and XROPEN[43]), its' constitutional constraints (XCONST[44]) and its' political participation (PARCOMP[45] and PARREG[46]) as presented in Figure 3.3.

Note that theses components capture the people's participation only with respect to participation in the political process, but no suffrage requirement is included. Hence "participation" as used by the Polity Project is not equal to the use of the term in the debates on measuring democracy. Rather the participation components evaluate the extent to which oppositional political activity is possible and regulated. The Polity Index is in fact a measure of political contestation rather than democracy - even if one embraces the minimalist democracy definition with contestation and participation.[47]

Figure 3.4 depicts the aggregation rule of Polity and the weighing scheme for Democ and Autoc. A country-year receives a score (scale weight) for each component. The points scored for Autoc are subtracted from the Democ index to obtain the Polity Index. Aside from a vague theory (explaining the "logic of institutionalized democracy and autocracy", Marshall et al. (2017a), p. 15 and p. 16), there is no justification given for the weighing and aggregation rule. Each of the component variables (XRCOMP, XROPEN, XCONST, PARCOMP, PARREG) is coded using three or more categories. However, not all of the categories are taken into account when calculating the democ/autoc and consequently the Polity Index. In a thorough examination of the Polity aggregation procedure Goertz (2006) shows that the Polity Index suffers from

[40] More information on the PolityIV Project is available at: `http://www.systemicpeace.org/polityproject.html`.

[41] Two types of datasets are offered: Polity IV Dataset in country-year format and the Polity IVd Dataset, where "d" denotes the country-date format.

[42] "Competitiveness refers to the extent that prevailing modes of advancement give subordinates equal opportunities to become superordinates", Marshall et al. (2017a), p. 21.

[43] "Recruitment of the chief executive is "open" to the extent that all the politically active population has an opportunity, in principle, to attain the position through a regularized process", Marshall et al. (2017a), p. 22.

[44] "(...) the extent of institutionalized constraints on the decisionmaking powers of chief executives, whether individuals or collectivities. Such limitations may be imposed by any "accountability groups." (...) The concern is therefore with the checks and balances between the various parts of the decision-making process. ", Marshall et al. (2017a), p. 24.

[45] "(...) the extent to which alternative preferences for policy and leadership can be pursued in the political arena.", Marshall et al. (2017a), p. 26.

[46] "Participation is regulated to the extent that there are binding rules on when, whether, and how political preferences are expressed.", Marshall et al. (2017a), p. 25.

[47] In the Appendix a factor analysis of the Polity Index' components is presented. It shows that the variation in the components can be explained by one latent factor - political contestation.

	$A = Autoc, A < 0; D = Democ$ $Polity = Democ + Autoc$	
Authority Coding:	Scale Weight	Counted in...
Competitiveness of Executive Recruitment (XRCOMP):		
(1) Selection	-2	A
(2) Transitional	+1	D
(3) Election	+2	D
Openness of Executive Recruitment (XROPEN):		
for DEMOC: only coded if XRCOMP is Election (3) or Transitional (2)		
for AUTOC: only coded if XRCOMP is coded Selection (1)		
(1) Closed	-1	A
(2) Dual/designation	-1	A
(3) Dual/election	+1	D
(4) Election	+1	D
Constraint on Chief Executive (XCONST):		
(1) Unlimited authority	-3	A
(2) Intermediate category	-2	A
(3) Slight to moderate limitations	-1	A
(4) Intermediate category	+1	D
(5) Substantial limitations	+2	D
(6) Intermediate category	+3	D
(7) Executive parity or subordination	+4	D
Competitiveness of Political Participation (PARCOMP):		
(0) Not Applicable	not counted in democ/autoc	
(1) Repressed	-2	A
(2) Suppressed	-1	A
(3) Factional	+1	D
(4) Transitional	+2	D
(5) Competitive	+3	D
Regulation of participation (PARREG):	AUTOC ONLY	
(1) Unregulated	not counted in democ/autoc	
(2) Multiple Identity	not counted in democ/autoc	
(3) Sectarian	-1	A
(4) Restricted	-2	A
(5) Regulated	not counted in democ/autoc	

Table 3.4: Aggregation Rule of the Polity Index. Source: Polity IV Dataset Users' Manual, Marshall et al. (2017a), p.15, 16 and 26

concept-measure inconsistency, i.e. the measure does not capture what the underlying concept of democracy very well.

Since the idea behind Polity was to quantify institutional frameworks it reached the limits of its domain in cases where there was no regular institutional setup. Years in which a country's central political authority is collapsed are considered as an interregnum period and coded "-77" on Democ, Autoc and Polity. Years in which central authority is taken over by foreign powers are considered interruption periods and coded "-66" for each index. During transition periods in which "new institutions are planned, legally constituted, and put into effect"[48] the indices receive a "-88" coding. Of 17,228 observations in the Dataset Version 2016, this is the case for 772 observations, i.e. around 4.5% of the observations. These three categories limit the use of the Polity Index in empirical research. To integrate these cases into the Polity scheme and make

[48]Marshall et al. (2017a), p. 19.

them accessible for quantitative studies the Polity IV Project introduced Polity2. It is defined as follows:

$$Polity2 := \begin{cases} \text{Polity Index}, & \text{if Polity Index} \in [-10, 10] \\ \text{missing value}, & \text{if Polity Index} = -66 \text{ (foreign interruption)} \\ 0, & \text{if Polity Index} = -77 \text{ (interregnum)} \\ \text{prorated across transition}, & \text{if Polity Index} = -88 \text{ (transition)} \end{cases}$$

Even though this definition enables the inclusion of transition and interregnum cases in time series and cross section models it creates a range of problems: First, the foreign interruption observations are still missing. In the Dataset Version 2016 there are 233 such cases, i.e. 1.3% of all observations. Depending on the goals of the research at hand, especially when examining democratic transtitions/autocratic backsliding or democracy and civil conflict, these missing cases are of particular interest. Second, a Polity2 value of "0" can occur in three different cases.[49] The most intuitive one is if a country's Autoc score equals its Democ score as it does for example in the case of Albania in 1996. The second reason why a country might be assigned a Polity2 value of "0" in a given year is a transition period. If a country is undergoing such a transition its Polity2 value is prorated across the time span of the transition. For example, if it has a transition year in 1990 and index values of "-2" in 1989 and "2" in 1991, then 1990 will be assigned a "0". Both of these codings are still somewhat suitable for empirical research since the differences between the Polity2 values remain meaningful. However, this is not true for the third case in which a country might be assigned a value of "0". In years in which a total collapse of central political authority occurs the country is assigned a Polity2 value of "0". The civil war in Afghanistan, 1992 - 1995, is an example for such a case of "anarchy". The meaning behind this particular "0" is rather different from the others and it renders differences between Polity2 values impossible to interpret. This "0-coding" hence affects the measurement scale: Polity2 is coded on a nominal (not even an ordinal!) scale limiting the usability of Polity2 in econometric models (at least if the "0-coding" is maintained).

In the Polity III Data (a predecessor of the Polity IV Dataset) some categories of the components on competition (PARCOMP) and regulation of political participation (PARREG) were defined with explicit reference to conflict.[50] Hence, countries experiencing severe civil conflict were highly unlikely to be classified as high/low democracies. Rather they were categorized as semi-democracies/anocracies leading to a number of studies[51] examining why semi-democracies seemed to be more prone to conflict than "pure" democracies/autocracies. Being aware of this point of criticism the Polity IV Project removed the explicit references from the definitions of the components.[52] However, even though the categories do not include the explicit reference anymore, they still capture the concept of a country undergoing conflict.[53] Due to the categorization of interregnum, interruption and transitions as described above and this coding problematic the aggregate Polity2 should not be used in research on civil conflict.

[49] Gleditsch and Ward criticised a similar point, namely that each Polity Index value can be achieved by a large number of different scores in the components and thus reflecting quite different polities in Gleditsch and Ward (1997).

[50] PARCOMP, categories (0 - unregulated) and (1 - factional) were coded with reference to civil war and violent conflict in the Polity IIId Data. Similarly, PARREG category (factional/restricted) exhibited the same problem, see Vreeland (2008), p. 406.

[51] Some of the most notable ones are Hegre (2001); Vreeland (2008); Fearon and Laitin (2003), pp. 84.

[52] See Marshall et al. (2017a), pp. 25.

[53] See for example the definition of PARCOMP's category (3 - factional competition): "Polities with parochial or ethnic-based political factions that regularly compete for political influence in order to promote particularist agendas and favor group members to the detriment of common, secular, or cross-cutting agendas.", Marshall et al. (2017a), p. 27.

One of the strongest advantages of the Polity IV Dataset is the availability of the disaggregate data.[54] This enables the breakdown of Polity2 into its components. Once the issues mentioned above (factionalism, missing interruption values etc.) are adequately adressed the components can be reassembled to form a measure of political contestation that (while still not continuous) can be employed in certain empirical models.

3.2.2 Freedom House

Freedom House is a US-based non-governmental organization dedicated to the promotion of freedom and democracy worldwide. Founded in 1941 in the midst of WWII it has significantly amplified it's sphere of operation and influence with the publication of annual "Freedom in the World" reports since 1973. The reports are based on annual surveys of global political rights and civil liberties. The survey data is available for 209 countries and territories from 1972 to 2016. With 209 countries covered FHI is the most inclusive of the three measures studied. The first reports and ratings were single-authored by Raymond Gastil, who argued in Inkeles (1991): "By working alone the author has not had to integrate the judgments of a variety of people. The hunches and impressions that are so important in a survey of this kind would be almost impossible to keep on the same wave lengths if one had an Asianist, Africanist, and Latin Americanist to satisfy before the ratings were finalized for each year".[55] While today the reports are produced by a team of "external analysts"[56], the checklist question framework introduced by Gastil and its inherent subjectivity remains. The FHI is not built upon any theoretical concept of democracy or freedom, rather it is a country comparison of an undefined concept of "freedom" based upon said "hunches and impressions". Bush (2017) shows that these impressions correspond to a large degree to the US elite's perception of other countries systems of governance and hence proposes taking FHI as a measure of such.[57] The remainder of this section will further illustrate this by providing a short overview of the components and aggregation procedure as well as the most prominent points of criticism.[58]

The Freedom in the World Index (FHI) evaluates the freedom concept along two dimensions: freedom of political rights and civil liberties. It is assembled in three steps. First, the Freedom House coders award from 0 (smallest) to 4 (greatest degree of freedom) points/scores to 27 questions. 12 of these questions regard the political rights dimension while the remaining 15 questions address the implementation of civil liberties.[59] Instead of four clear cut possible answer categories (0 - 4 points are rewarded) for each of these questions a number of subquestions are given to clarify the concept. The disaggregate data is not publicly available, hence there is no way of knowing or replicating how a country achieved a certain score or of testing the implications of the aggregation rule.[60]

In a second step, depending on the sum of scores obtained in the political rights and civil liberties components a rating is assigned. The scores to rating conversion for both components is presented in Table 3.5 for the Political Rights Index.[61] The conversion for the Civil Liberties

[54] Also noted by Munck and Verkuilen (2002), p.20.
[55] See Inkeles (1991), p. 22.
[56] See Freedom House (2017a).
[57] See Bush (2017), p.725.
[58] Since these points of criticism are plenty but the number of pages is not, the critiques are discussed briefly and the interested reader is asked to refer to the corresponding articles for further information.
[59] The Appendix provides a table displaying the concrete questions and respective scores.
[60] This has been frequently criticised, for example by Munck and Verkuilen (2002), p. 25; Munck (2009), p. 29; Cheibub et al. (2010), p. 75.
[61] Note, that while Table 3.5 depicts 0 as the minimum score a country can be given "It is possible for a country's or territory's total political rights score to be less than zero (between -1 and -4) if it receives mostly or all zeros for

Total Scores	Political Rights Rating	
36 - 40	1	Greatest range of political rights implemented
30 - 35	2	Intermediate category (between 1 and 3)
24 - 29	3	Countries and territories with a rating of 3, 4, or 5 either moderately protect
18 - 23	4	almost all political rights or strongly protect some political rights while
12 - 17	5	neglecting others. The same factors that undermine freedom in countries with a rating of 2 may also weaken political rights in those with a rating of 3, 4, or 5, but to a greater extent at each successive rating.
6 - 11	6	Intermediate category (between 5 and 7)
0 - 5	7	Few or no political rights (...) sometimes in combination with civil war

Table 3.5: Score to rating conversion for the Political Rights Index, Source: Freedom House (2017b)

is carried out analogously. It is worth noting, that (due to the high number of possible ways of obtaining a certain score rating) each of the seven categories captures numerous countries with very distinct political rights and civil liberties.

In a third step the political rights and civil liberties ratings are averaged to form the freedom rating. It ranges between 1 and 7. This rating is used to categorize the countries into three groups: Free (rating between 1 - 2.5), Partly Free (rating between 3 - 5) and Not Free (rating between 5 - 7). As mentioned above, each of the seven categories contains a wide range of countries with politically very different environments. Further categorizing them into three groups inherently continues disguising this variation.[62]

One of the most criticised aspects of the Freedom House data is the compilation of components by means of checklist questions without a theoretical justification. This is problematic in several ways: First, the components are not ordered by level of abstraction and the relationship between the components is not considered. This gives rise to the problem of conflation[63] as criticised by Munck and Verkuilen (2002), p. 14.; Coppedge et al. (2011), rightfully observe that "the high inter-correlations of the Freedom House indicators coupled with their ambiguous coding procedures suggest that these components may not be entirely independent of one another".[64] Second, as mentioned above there are no clear cut answers for the checklist questions. Since the questions are formulated in a way as to capture highly subjective features (e.g. "Are the electoral laws and framework fair?") the lack of clear answers transmits this problem of subjectivity further into the data. This was, for example, criticised by Cheibub et al. (2010), p. 75.

each of the 10 political rights questions and it receives a sufficiently negative score for political rights discretionary question B. In such a case, it would still receive a final political rights rating of 7", Freedom House (2017b). The discretionary political rights questions (see Appendix) reduce the political rights scores without a clear rule of application thus contributing to a further subjective bias in the data.

[62] "For example, those at the lowest end of the Free category (2 in political rights and 3 in civil liberties, or 3 in political rights and 2 in civil liberties) differ from those at the upper end of the Free group (1 for both political rights and civil liberties). Also, a designation of Free does not mean that a country or territory enjoys perfect freedom or lacks serious problems, only that it enjoys comparatively more freedom than those rated Partly Free or Not Free (and some others rated Free)", Freedom House (2017b).

[63] See section *Definition of democracy*.

[64] See Coppedge et al. (2011), p. 251.

Another point of concern is the inappropriate aggregation rule of addition - equal weighing. Asigning equal weights to each question asked/concept contained is disputable in light of their content. This aggregation rule does not capture the complementarity of the concepts participation and contestation. Furthermore, it does not view them necessary conditions for a democracy.[65] It is also noteworthy that the measurement scale is neither continuous nor ordinal, it is at best categorical.[66]

There is also concern with respect to the usage of the FHI in a time series context. A number of studies have exposed an ideological bias in the Freedom House Data: Bollen and Paxton (2000), p. 77, for example find evidence for a systematic downrating of Marxist-Leninist countries, especially in Freedom House's early years and abating around the mid 1980s. In addition, in some years the coding rules are altered from one year to the next and previous years are not updated. Cheibub et al. (2010), p. 75, and Munck (2009), p. 148 (footnote 15) observe that hence, the use of Freedom House Data in a time series context is hardly justifiable.

The problem of subjectivity is also inherent in the coding of the time series. Freedom House states: "the scores from the previous edition are used as a benchmark for the current year under review. A score is typically changed only if there has been a real-world development during the year that warrants a decline or improvement (e.g., a crackdown on the media, the country's first free and fair elections), though gradual changes in conditions, in the absence of a signal event, are occasionally registered in the scores".[67] This benchmarking can potentially lead to the transmission of a subjective coding bias over long periods of time.

In conclusion, Freedom House Data should be used in statistical analyses with extreme caution. The Dataset "by Freedom House (...) exemplifys problems in all three areas of conceptualization, measurement, and aggregation", Munck and Verkuilen (2002), p. 28.

3.2.3 Varieties of Democracy (V-Dem) Dataset

The Varieties of Democracy (V-Dem) Dataset is assembled by a cooperation of over 50 scholars from all over the world, co-hosted by the Department of Political Science at the University of Gothenburg, Sweden, and the Kellogg Institute at the University of Notre Dame, USA. Several of the scholars involved in the evolution of the V-Dem Dataset have contibuted to the literature on democracy measurement long before the V-Dem came to life (for example Michael Coppedge, Carl Henrik Knutsen, Jan Teorell or Pamela Paxton to name a few). They were well aware of the ongoing debate about the definition, scaling and aggregation of existing democracy measures. As a result the V-Dem Dataset provides answers to several of the problems discussed in the first section and is probably the most stringent and transparent Dataset on democracy available today.

A first version of the dataset was introduced in Lindberg et al. (2014). The most recent version of data (the one used in this paper) is Version 7.1, Coppedge et al. (2017d) The data is available in country-year as well as in country-date format.

Numerous country experts, coordinators and research assistants code basic variables capturing distinct features of a democratic system.[68] The variables are distinguished by type: type A vari-

[65] As criticised by Munck and Verkuilen (2002) p. 25; Munck (2009), p. 50 - 51.
[66] See Cheibub et al. (2010), p. 75.
[67] See Freedom House (2017b).
[68] For more information on the variable types and coding procedure please refer to Coppedge et al. (2017a) pp. 36.

ables are "based on extant sources and (...) factual in nature"[69] while type B and C variables are coded by country coordinators and experts. Type B variables are more factual items than C and do not contain as much judgment. Type C variables on the other hand require a greater amount of judgment and country specific knowledge, such as language and state of affairs. The question of how to aggregate variables coded by multiple experts into a single "best estimate" for each variable is quite important in this context (for example when some of the expert coded questions are coded on an ordinal scale). Most of the (C) variables are based on questions with answers on an ordinal scale and are thus aggregated across coders using Bayesian item response theory models.[70] Once the basic variables are aggregated into indices they are assigned type D.[71]

One of the main qualities distinguishing the V-Dem Dataset from others is their definition and conceptualization of democracy: V-Dem acknowledges the fact that a democracy measure's validity/its concept-measure consistency hinges on the proper definition of the underlying concept. Given the multi-faceted nature of the concept of democracy, they provide disaggregate data. This gives empirical scholars the opportunity to construct democracy measures based upon concepts defined as needed in particular research frameworks. As a potential starting point, they propose considering democracy as a multidimensional concept consisting of the following five distinct dimensions:

1. The electoral dimension[72]
2. The participatory dimension[73]
3. The egalitarian dimension[74]
4. The deliberative dimension[75]
5. The liberal dimension[76]

In addition, V-Dem also recognizes the importance of an aggregation procedure reflecting the theoretical relationships between the concept's attributes for concept-measure consistency. Low level indices (D-type) in the V-Dem Dataset are combined into several mid-level indices, which in turn are then aggregated into high level indices reflecting these five dimensions of democracy. In this aggregation process issues of complementarity/substitutability or family resemblance between these dimensions are addressed. In the V-Dem definition of democracy, the electoral dimension is circled out as the core element without which no country shall be labelled democratic.[77] Hence, the Electoral Democracy Index is combined with the high level indices of

[69] See Coppedge et al. (2017a), p. 36.

[70] The measurement models are described in detail in Coppedge et al. (2017e), pp. 29. Marquardt and Pemstein (2018), further discuss item response theory models and compare their performance.

[71] More detailed information on the variable types can be found in Coppedge et al. (2017e), p. 17 - 18.

[72] "(...) embodies the core value of making rulers responsive to citizens through competition for the approval of a broad electorate during periodic elections", Lindberg et al. (2014), p. 160.

[73] "(...) embodies the values of direct rule and active participation by citizens in all political processes; it emphasizes nonelectoral forms of political participation such as through civil society organizations and mechanisms of direct democracy.", Lindberg et al. (2014), p. 160.

[74] "(...) holds that material and immaterial inequalities inhibit the actual exercise of formal rights and liberties; hence a more equal distribution of resources, education, and health across socioeconomic groups should enhance political equality.", Lindberg et al. (2014), p. 161.

[75] "(...) enshrines the core value that political decisions in pursuit of the public good should be informed by respectful and reasonable dialogue at all levels rather than by emotional appeals, solidary attachments, parochial interests, or coercion.", Lindberg et al. (2014), p. 160.

[76] "(...) embodies the intrinsic value of protecting individual and minority rights against a potential "tyranny of the majority." This is achieved through constitutionally protected civil liberties, strong rule of law, and effective checks and balances that limit the use of executive power." , Lindberg et al. (2014), p. 160.

[77] See Lindberg et al. (2014), p. 161.

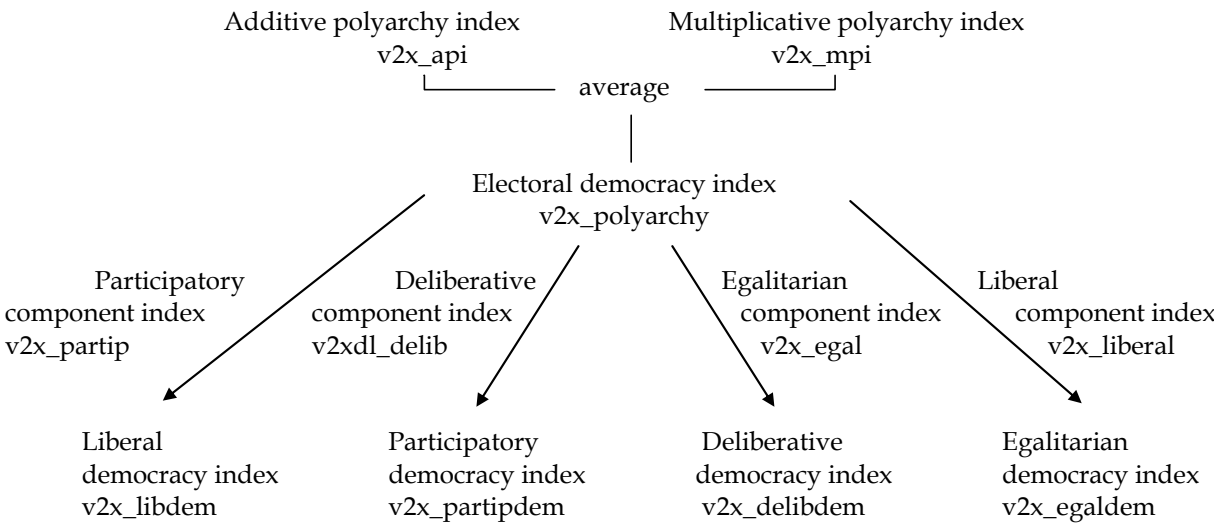


Figure 3.1: Aggregation process for high level V-Dem Indices

dimensions 2 - 5 to create four indices of democracy.[78] These five indices are considered to embody the "varieties of democracy" (V-Dem).

Detailed information regarding V-Dem disaggregate data is given in the V-Dem Codebook, Coppedge et al. (2017a). The basic A,B,C-type variables are aggregated into low level indices (D-type), for example by "adding a denominator (e.g., per capita), by creating a cumulative scale (total number of...), or by aggregating larger concepts (e.g., components or indices of democracy)", Coppedge et al. (2017a), p. 36.

The Electoral Democracy Index, Polyarchy, reflects the minimalist democracy definition of contestation and participation.[79] To construct it, five low level indices reflecting freedom of expression, association, suffrage and free and fair elections are aggregated into two mid level indices of electoral democracy: the additive and the multiplicative polyarchy index. The additive index is constructed by weighted addition of the five low level indices and reflects the substitutability of democracy's attributes. The multiplicative polyarchy index on the other hand is constructed by multiplying the five low level indices and captures the complementarity of the five concepts (a very low score on one of the components will lower its' overall multiplicative polyarchy index). The Electoral Democracy Index, joins complementarity and substitutability by averaging the additive and multiplicative polyarchy indices. The aggregation process for the other four high level democracy indices is analogous. It is displayed in Figure 3.1.

The arrows represent an aggregation procedure which is averaging multiplication (complementarity) and addition (substitutability) of the two components:

$$\text{democracy index} = \tfrac{1}{4} \cdot (\text{polyarchy}^{1.6} + \text{component index}) + \tfrac{1}{2} \cdot \text{polyarchy}^{1.6} \cdot \text{component index}$$

Polyarchy influences each high level democracy index to the power of 1.6.[80] Setting a higher rate of influence for Polyarchy than any of the other component indices underlines the impor-

[78] The Liberal Democracy Index, v2x_libdem, the Participatory Democracy Index, v2x_partipdem, the Deliberative Democracy Index, v2x_delibdem as well as the Egalitarian Democracy Index, v2x_egaldem.

[79] The Appendix contains a figure displaying the components and aggregation rule for Polyarchy.

[80] The derivation of this rate is explained in the Appendix as well as in Coppedge et al. (2017e), p.10.

tance of the electoral democracy principle. Contestation and participation should be satisfied to a certain degree before further aspects of democracy can be employed to distinguish between higher levels of democracy.

The V-Dem Indices are continous and range between [0,1]. Since in some instances ordinal measures might be needed, the V-Dem Project also offers the main indices as ordinal variables with 3, 4 or 5 categories respectively. The classification rules for the ordinal indices are also provided, Coppedge et al. (2017a), pp. 268.

In addition to the democracy indices the V-Dem Dataset provides measures of uncertianty for each of the indices. For each of the indices the standard deviation (index suffixed _sd) as well as the "interval in which the measurement model places 68 percent of the probability mass for each country-year score"[81] (denoted by index suffixed _codelow and _codehigh) are given. These highest posterior density intervals can be seen as an indication of the skewedness of the underlying distribution: if the distances between the point estimate and the upper and lower bound are not equal, the underlying posterior distribution is skewed.

The dataset is very transparent. Even the project manager, responsible for crafting a specific variable, or the compiler is listed in the data. The number of experts coding a variable is also given (variabled suffixed with _nr). The default number of coders for the period 1900-2012 is 5 or more.

The project documentation is extensive. Aside from the Coppedge et al. (2017a), the project for example offers documentation on methodology, Coppedge et al. (2017e), on the units of observation, Coppedge et al. (2017c), as well as on the project's setup, Coppedge et al. (2017f).

To ensure consistency of codings accross countries and over time bridge and lateral codings are employed.[82] Bridge coding refers to one coder coding a certain questions for multiple countries over the entire time series while lateral coding refers to the coding of a variable across all countries in a single year. Acknowledging the fundamental role of the country experts V-Dem provides information on country expert selection processes.[83] To avoid repressions the contry experts remain anonymus.[84]

An important feature of the V-Dem data to note is its sensitivity due to the Bayesian item response theory models. With addition of "new information", for example through the addition of a new bridge or lateral coder or the addition of a new year the point estimates for the indices change from each dataset version to the next. As a result country scores differ between the versions. While this might seem odd at first, this sensitivity to new information is one of the biggest strengths of the data: While Polity2 is struggling with the lack of variation in high democracies and low autocracies (as mentioned in section Polity Index) the above mentioned sensitivity of the V-Dem data will facilitate its adaption to new international constellations without loss of variation. Furthermore, it highlights that the country scores are point estimates with inherent uncertainty. This is a crucial point to keep in mind especially when creating rankings based on democracy measures.[85]

One minor issue to consider with the V-Dem dataset is that their distinction between de facto and de jure aspects of a polity (and hence their influence in the democracy measures) is not

[81] Coppedge et al. (2017a), p. 37.
[82] See Coppedge et al. (2017e), pp. 25.
[83] See section on country expert recruitment in Coppedge et al. (2017e), p.18.
[84] See Coppedge et al. (2017e), p.21.
[85] See Høyland et al. (2012) for a thorough discussion of the pitfalls of disregarding uncertainty.

discussed or illustrated. Polyarchy, for example, contains a suffrage requirement and thus a measure of institutionalized enfranchisement (de jure participation). However, it is questionable how far actual (de facto) participation is captured. The data introduced in Vanhanen (2000) as well the Scalar Index of Polities by Gates et al. (2006), for example, measure participation as the percentage of the population which actually voted in the most recent parliamentary or presidential election (or both).[86] Considering voter turnout as a sole measure of participation can produce misleading results, for example if voting is mandatory. To capture a de facto aspect of participation Polyarchy includes an indicator on electoral violence.[87] A discussion of the inclusion of de jure/de facto criteria in the definition of democracy would be desirable.

Nevertheless, with the public availability of disaggregate and aggregate data, theoretical justification for the detailed aggregation rule and comprehensive spatial and temporal coverage, the V-Dem dataset provides the most well-documented and well-grounded collection of democracy measures available today.

3.3 Comparison

The previous section outlined the diverse definitions and aggregation procedures embedded in each of the democracy measures. In addition, the three datasets vary considerably in terms of countries and years covered. These differences in definition, in availability of disaggregate data and country-years coded by each index are substantial and often prove decisive with respect to the question which index to use in which framework. To assess dissimilarities in coding between the three indices the sample is restricted to those observations available in all datasets for the remainder of this section. This sample will be referred to as the trunk dataset.

Note that while the reduction to a trunk dataset is necessary it is also a very harsh restriction and a considerable amount of information is disregarded. For example, restricting the sample to countries included in the Freedom House data limits the timeline to post-1972 years. However, Polity2 and V-Dem Data can be compared using data from 1900 on. It becomes even more drastic when the observations left out are chosen by attribute (and not by year): In the Polity2 coding scheme observations during which a country is occupied by foreign powers constitute missing values (which is not the case for Polyarchy and FHI). Thus, these cases are not part of the trunk dataset.

How does this restriction affect similarities between the measures? Comparing pairwise correlations[88] to the correlations obtained using the trunk dataset (see Table 3.6) shows that for the observations in the trunk dataset (top right) the indices' correlation is larger or (almost) equal to the pairwise comparison (displayed in the bottom left part).

Coppedge et al. (2011)[89] point out that high intercorrelations between the democracy indices are at least partly due to observations which are "perfectly" democratic/autocratic. The trunk dataset contains 6,546 observations for 167 countries from 1972 to 2015. In comparison to the period 1900 to 1971 the number of "perfect" democracies drastically increased after 1972, which could explain the higher intercorrelations in the trunk data. As displayed in Table 3.7 around 21.2% of the observations in the trunk dataset are coded as perfect democracies/autocracies

[86] See Vanhanen (2000), p. 253, and Gates et al. (2006), p. 897.

[87] Election other electoral violence, v2elpeace, see Coppedge et al. (2017a), p. 97.

[88] Note, that the V-Dem Data starts in 1900. When computing pairwise correlation coefficients thus the number of observations is much higher than when doing so for the trunk dataset.

[89] See Coppedge et al. (2011), p. 252.

	Polity2	Polyarchy	FHI
Polity2	1	0.9083	0.8889
		(6,546)	(6,546)
Polyarchy	0.8661	1	0.9219
	(11,781)		(6,546)
FHI	0.8892	0.9200	1
	(6,580)	(6,902)	

Table 3.6: Pairwise correlation coefficients, bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis below.

Number of perfect...	1900-1971	1972 - 2015
... autocracies	265	194
... democracies	1,026	1,194
Total	1,291	1,388

Table 3.7: Number of perfect democracies/autocracies (+/-10 in the Polity2 coding scheme) in the period 1900 - 1971 and the trunk dataset (1972 - 2015)

(+/- 10 on the Polity Scale). When removing them from the sample the correlations decrease as displayed in Table 3.8. Limiting the dataset to observations available for all three indices thus implies obtaining a sample with a high fraction of perfect democracies/autocracies, which in turn contributes to an intercorrelation between indices which is larger or (almost) equal to the pairwise comparisons.

3.3.1 Summary statistics

Table 3.9 displays summary statistics for the democracy indices discussed above. Since it represents the minimalist democracy definition of contestation and participation the Polyarchy Index[90] was chosen from the V-Dem indices.[91] For facilitated comparability the FHI was reversed and both, FHI and Polity2 were normalized between 0 and 1.[92] Due to different geographical and temporal scopes covered by each index the number of observations for which the summary statistics are computed are quite different.[93] The summary statistics for all observations available per index are displayed on the top part of Table 3.9. The summary statistics for the trunk dataset are displayed in the lower part of Table 3.9.[94] The distributions of all three measures are u-shaped,[95] with peaks (particularly pronouced for Polity2 and FHI) at the extremes.[96] As discussed above 18.2% of the observations in the trunk dataset are coded as "perfect" democracies by Polity2. This high number of perfect democracies is also reflected by a Polity2 median

[90] Polyarchy was used exactly how it is provided in the data, i.e. it was not rescaled or normalized.

[91] Summary statistics for the other V-Dem democracy indices are given in the Appendix.

[92]
$$\text{Normalized Polity2} = \frac{Polity2 + 11}{21}, \text{ reversed and normalized FHI} = \frac{7 - FHI}{6}$$

[93] More information on the different geographical and temporal scopes covered by Polity2 and Polyarchy can be found in Boese and Kamin (2018c).

[94] The Appendix includes a list of countries and their respective years coded in the trunk dataset.

[95] Several studies, such as Goertz (2006), discuss how colinearity of the indicators forming the democracy indices can contribute to to this u-shape.

[96] The Appendix provides a histogram as well as a table with percentiles for the three democracy measures in the trunk data giving further insight into their empirical distribution.

	Polity2	Polyarchy	FHI
Polity2	1	0.8648	0.8423
		(5,158)	(5,158)
Polyarchy	0.8002	1	0.8710
	(9,169)		(5,158)
FHI	0.8425	0.8710	1
	(5,173)	(5,158)	

Table 3.8: Pairwise correlation coefficients for observations which are not coded "perfect" democracies/ autocracies (10/-10) in Polity2.
Bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis

	Summary statistics for...					
	...all observations available per index					
Variable	Obs.	Mean	Median	St. D.	Min	Max
Polity2	16,992	0.4724	0.35	0.3536	0	1
Polyarchy	17,036	0.3179	0.2055	0.2788	0.0086	0.9471
FHI	6,936	0.5071	0.5	0.3375	0	1
	...the trunk dataset					
Variable	Obs.	Mean	Median	St. D.	Min	Max
Polity2	6,546	0.5689	0.7	0.3647	0	1
Polyarchy	6,546	0.4545	0.4061	0.2868	0.0140	0.9471
FHI	6,546	0.5053	0.5	0.3362	0	1

Table 3.9: Summary statistics for the democracy indices

which is substantially larger than Polity2's mean. The distributions of FHI and Polyarchy are more even. However, FHI considers around 14% of observations in the trunk dataset as perfect democracies and around 9% as perfect autocracies while no observation is coded as either by the V-Dem Project. Polity2 and FHI capture very little variation in highly democratic/autocratic systems (the US, for example, is coded as a perfect democracy from 1972 to 2015) as opposed to Polyarchy.[97] This has an important implication for the choice of measure in frameworks in which highly democratic/autocratic countries are examined. The FHI also has a comparatively high number (around 8.5% of the observations) of perfect autocracies, i.e. observations coded as 0. In addition, the high and low numbers of perfect democracies/autocracies have an noteworthy theoretical implication for the future development of the respective countries' time series: for those countries the system of governance is not able to improve/deteriorate.

3.3.2 Examining the differences

The following section examines the differences in coding between the three democracy measures and their implication for the country ranking within each index. The section closes with a country study exemplifying the differences in coding variation, in ranking as well as in disaggregation possibilities.

[97] The Appendix includes an exemplary discussion of the within and between country variation in the Polity2 data.

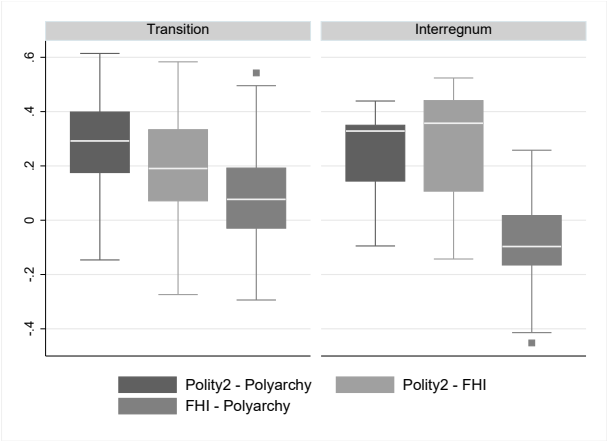


Figure 3.2: Boxplot of differences between index values for cases coded as transitions or interregnum in the Polity Scheme

Difference with respect to attributes of the democracy definition

Polity2 assigns a value of 0[98] for observations in which a country undergoes a period of interregnum/anarchy and prorates the respective country's democracy value over the length of a transition period (see Section *Polity Index*).

The boxplot in Figure 3.2 shows that Polyarchy and FHI are both much lower than Polity2 (i.e. the interquartile range of the differences is strictly positive) for countries in anarchy and transition periods. Interestingly, at the same time the difference in values between FHI and Polyarchy is low for these cases (median differences of around $+/-0.1$), while the differences between each of them and the Polity2 coding are substantial. Countries undergoing periods of anarchy or transition appear to be systematically overrated by Polity2. In research frameworks in which such countries play an important role FHI or V-Dem indices should therefore be preferred over Polity2.

As mentioned in Section Polity Index the original aim of the Polity data was to capture a country's formal institutional setup. This led to the Polity Index reaching the boundaries of its domain in cases where there were no regular institutions, i.e. in cases of interregnum, transition or occupation. Even though V-Dem includes several aspects beyond the formal institutions the absence of a regular institutional framework in post WWII-Germany seems to be posing difficulties for V-Dem as well: Germany is not coded between 1946 - 48. Since this is the only such case the recommendation to prefer V-Dem indices or FHI whenever a share of the relevant observations falls in categories discussed above remains valid.

Difference in coding by level of democracy

The high intercorrelations between the democracy measures discussed above show that there is a strong relationship between them. This is to be expected, as the three indices are supposed to measure the same thing - democracy. But, the correlations do not provide information on the "agreement" between the indices. Agreement can be thought of as the identity line, when two indices are plotted with respect to each other. If all observations are close to the identity line, the indices "agree" and the differences between them are close to zero. A case with high correlation, but limited agreement occurs, for example, if one index were to consistently code

[98] A Polity2 value of 0 translates into $11/21 \approx 0.52$ on the normalized polity scale (between 0 and 1).

each country half as democratic as the other.

In the following, the general agreement between the indices is examined and systematic differences are explored. Let's assume there was a "hypothetical democracy scale" along which all countries could be sorted. Does one index code highly democratic/autocratic countries (on the top/bottom of the hypothetical scale) systematically different from another index? The "hypothetical democracy scale" is, of course, unknown (and given the different definitions underlying the democracy measures it is strictly hypothetical). However, assuming the three democracy measures are three ways of capturing very similar concepts, we can use the average of the three measures as a proxy for the hypothetical scale. Differences between the index pairs are plotted against this average, i.e. the "hypothetical democracy scale" in Figure 3.3. The solid line at y=0 marks the observations for which there is perfect agreement between the indices (i.e. where the difference between them equals zero). The dashed lines mark the 95th-percentile and the 5th-percentile, hence, 90% of the observations fall between the top and bottom dashed line. The black line represents a linear fit (pooled OLS) of the differences in democracy measures on "hypothetical democracy".[99]

For all indices the differences get closer to zero towards the democratic/autocratic extreme. The plot for Polity2 and V-Dem Polyarchy Index provides useful insight on the agreement between the two indices: The fitted line is positive and increasing. In a large number of cases Polity2 rates countries more democratic than the Polyarchy Index. For the intermediate 90% of observations the difference between the two index values tends to be positive. In fact, for countries in the upper half of the democracy scale, there is very little agreement: The difference in index values becomes larger as the level of "hypothetical democracy" increases. The vast majority of countries located between 0.55 and 0.8 on the hypothetical democracy scale is coded strictly more democratic by Polity2 than by Polyarchy Index. For countries at the very low end (until 0.1) of the democracy scale, on the other hand, the two indices seem to agree to some extent: differences are symmetrically distributed close to 0. There are comparatively few observations for which Polyarchy is significanly larger than Polity2 (those in the 5-th percentile). They occur for countries on the lower half (between 0.1 and 0.7) of the hypothetical democracy scale.

The plot for FHI and Polyarchy also displays an increase in difference between the two measures with rising hypothetical democracy values. The fitted line is increasing, but it is negative for hypothetical democracy values below 0.16. Countries on the lower end of the hypothetical democracy scale are coded slightly more democratic by Polyarchy than by FHI. However, this is reversed for countries above the 0.18 threshold: for those FHI tends to be slightly larger than Polyarchy and increasingly so with rising levels of democracy. Large differences (as measured by the 10% of observations for which the absolute differences in democracy values are particularly high) occur for countries anywhere on the democratic scale except the end points.

Polity2 and FHI's agreement is almost consistent across hypothetical democracy values: The fitted line is nearly horizontal, but positive. Polity2 tends to code countries slightly more democratic anywhere on the democracy scale. Examining the large differences only, FHI tends to be larger than Polity2 only for countries in the lower half of the democratic scale.

The assessment of dis-/agreements between democracy measures concludes with testing whether there are significant differences in coding between the indices: is the mean difference between the indices is different from zero (H0)? The results are displayed in Table 3.10. They are in line with the findings from Figure 3.3 above. The mean differences are positive and significantly different from zero. On average Polity2 assigns the highest democracy values, followed by FHI.

[99] Detailed regression results are provided in the Appendix.

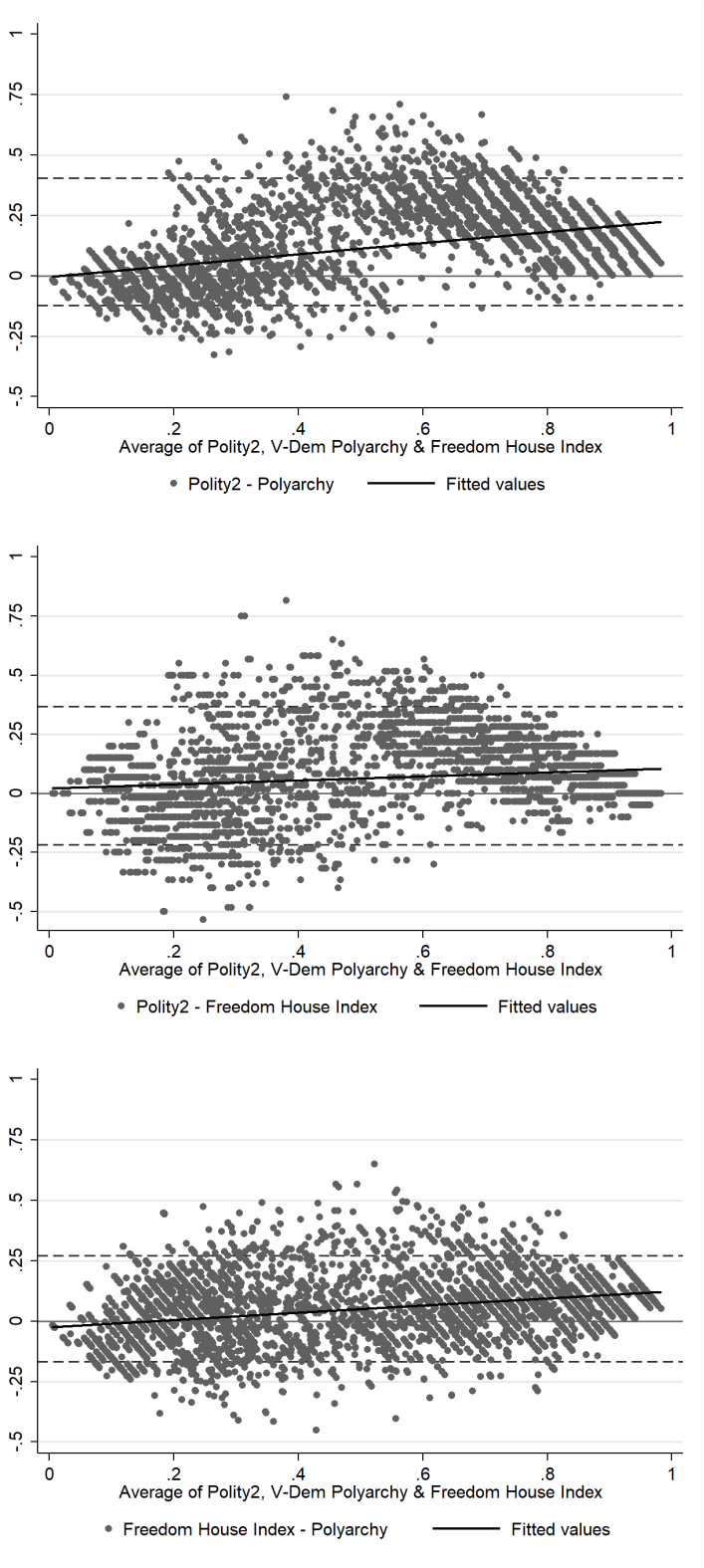


Figure 3.3: Pairwise differences between indices plotted over average democracy levels, horizontal lines at 0, the 95th-percentile and the 5th-percentile

Difference between...			Mean
Polity2	-	Polyarchy	0.1144***
Polity2	-	FHI	0.0636***
FHI	-	Polyarchy	0.0508***

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.10$

Table 3.10: Results of ttest of mean differences between indices

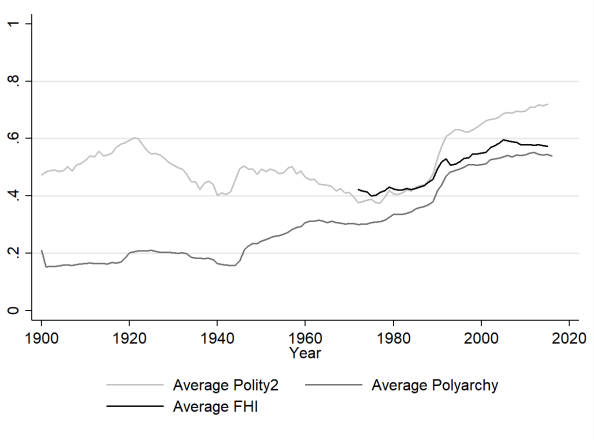


Figure 3.4: World averages of Polyarchy, Polity2 and FHI

Polyarchy on average assigns the lowest democracy values. Due to Polyarchy's definition this comes as no surprise: as discussed above it captures the minimalist democracy definition of contestation and participation.

Difference by region

To examine the results from the last Section world averages from 1900 - 2016 are plotted in Figure 3.4. Recall, that the Polyarchy Index captures the minimalist democracy definition of contestation and participation. When examining the world averages this becomes obvious as it is - on average - quite austere, i.e. lower than FHI or Polity2, in its coding. Interestingly, the world averages show that FHI values fluctuate between Polity2 and Polyarchy only from 1990 onwards. Before that the FHI average is above the Polity2 one.

Figure 3.5 breaks up the dataset by geopolitical region[100] and displays regional averages for each of the democracy measures. For most regions the Polity2 values are larger or equal (almost equidistant) to the Polyarchy values while the FHI mostly "fluctuates" between the other two. The change in coding between FHI and Polity2 averages noted in Figure 3.4 is particularly visible in the following regions: In Latin America & Caribbean, Middle East & North Africa as well as East and South Asia the Polity2 average is below the FHI average prior to 1990, but above it after. In addition, when examining the regional averages for Eastern Europe & Central Asia Polity2 and Polyarchy Index show a high level of aggrement for the observations before 1990. The Freedom House Index on the other hand codes the Eastern European observations much less democratic than the other two indices between 1972 - 1990. This might be a reflection of the Freedom House Index' early ideological bias discussed in Section *Freedom House*. It could also

[100] A detailed list of countries and their respective regions can be found in the Appendix.

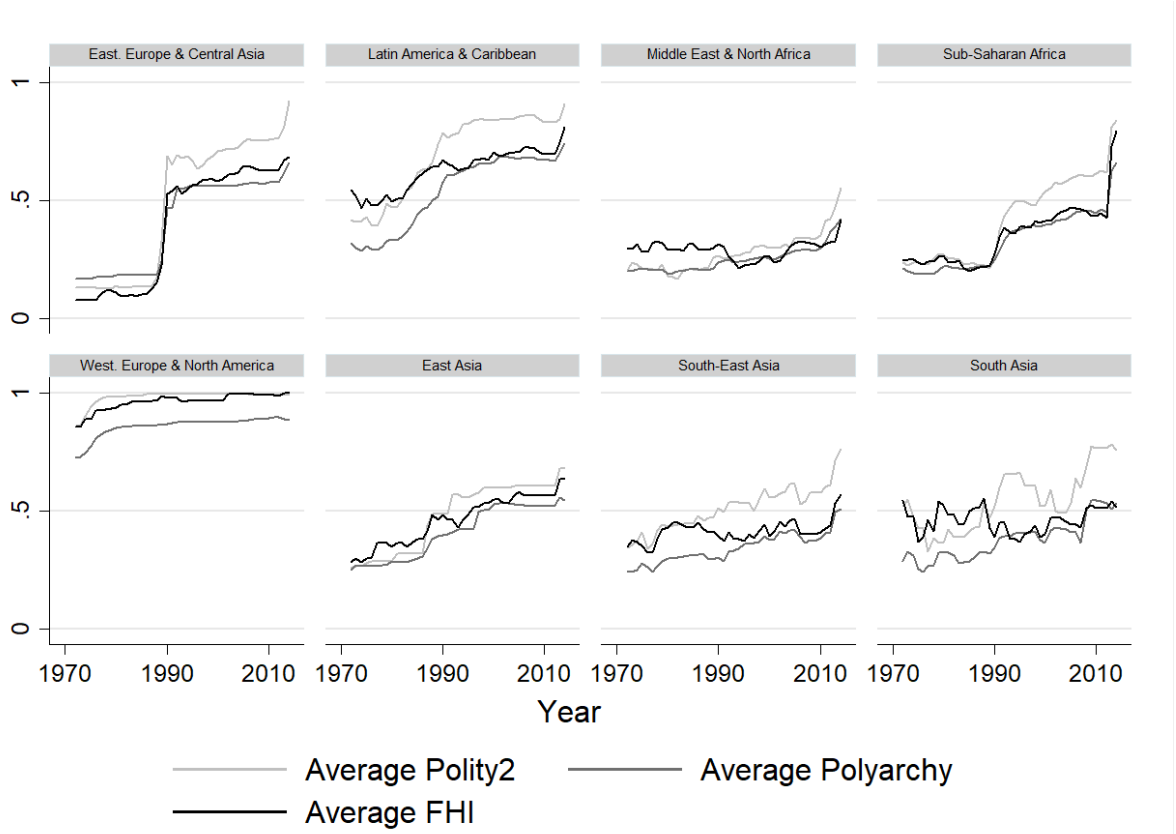


Figure 3.5: Regional averages of the three democracy measures

indicate Polyarchy values which are too high. It is the only timespan and region in which Polyarchy is higher than both, Polity2 and FHI. In the following Section, the coding of the Eastern European & Central Asian countries will be examined in more detail to shed light on this finding.

Eastern Europe & Central Asia

The regional comparison in the Figure 3.5 provides some insight into the "austereness" of each democracy index. On average Polity2 assigns the highest and Polyarchy the lowest values while the FHI fluctuates between the other two. The only time and region in which there seems to be a systematic deviation from this scheme is in Eastern Europe & Central Asia (EECA) before 1990. There are 31 countries in the region.[101] Albania, Bulgaria, Mongolia and Romania provide good examples of the coding phenomenon discussed above. Hence, their coding will be studied in more detail in the remainder of this section. The respective democracy values assigned by Polity2, FHI and Polyarchy are displayed in Figure 3.6.[102] To understand why Polyarchy codes each country comparatively high Polyarchy is broken up into its components in Figure 3.7. The share of population with suffrage as well as the Elected officials index[103] are coded with the highest possible value over the entire time span. This strongly contributes to the high Polyarchy

[101] A list of these countries is provided in the Appendix.

[102] The Appendix contains a Figure displaying the codings for all other countries in the EECA region. See Figure B.5.

[103] The elected officials index, v2x_elecoff addresses the question "Is the chief executive and legislature appointed through popular elections?", Coppedge et al. (2017a), p. 59.

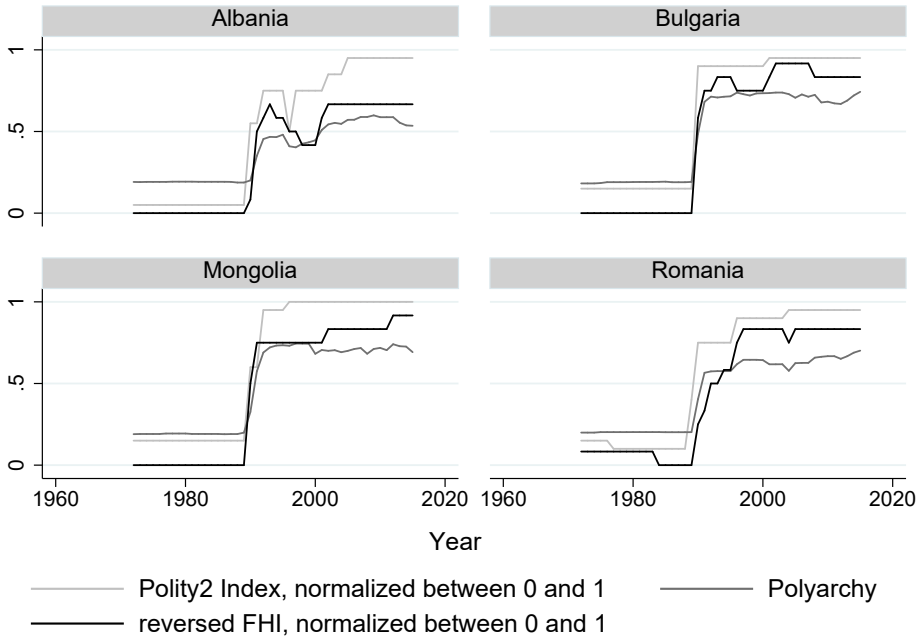


Figure 3.6: Democracy coding by Polity2, FHI and Polyarchy for Albania, Bulgaria, Mongolia and Romania

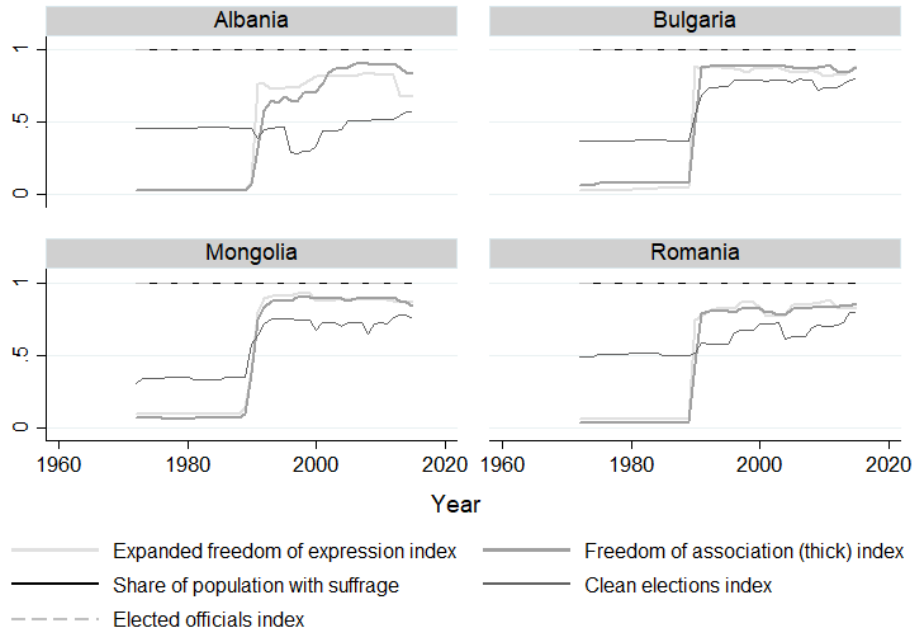


Figure 3.7: Components of Polyarchy for Albania, Bulgaria, Mongolia and Romania

values in these countries prior to 1990. It is a central difference between the three indices and their underlying definitions of democracy. As discussed above, Polity2 and FHI[104] do not include any components capturing suffrage requirements. Polity2's categories XROPEN ("Recruitment of the chief executive is "open" to the extent that all the politically active population has an opportunity, in principle, to attain the position through a regularized process", Marshall et al. (2017a), p. 22.) and XRCOMP ("Competitiveness refers to the extent that prevailing modes of advancement give subordinates equal opportunities to become superordinates", Marshall et al. (2017a), p. 21.) are closest in meaning to V-Dem's elected officials index. Here a lack of detail and organization in theoretical concept of the Polity2 scheme becomes apparent: the very definition of XROPEN mixes several important attributes, namely the regularized electoral process (in V-Dem captured by the elected officials index), the extent to which the politically active population is *de jure* allowed to participate (i.e. the share of population with suffrage) as well as the actual *de facto* opportunities the population has to influence the electoral process (in V-Dem this is covered, for example, through the clean elections index). The Polity2 categories are not distinguishing between different attributes of the underlying concepts of the de jure electoral process and the de facto as well as de jure participation of the population.

Section B.8 of the Appendix compares the country ranking of each of the three democracy indices. It shows that the countries are ranked almost consistently across all regions except for EECA prior to 1990. The difference in the democracy definition covered (by the inclusion of suffrage) is a major driver behind the comparatively high values of Polyarchy in that time and region. It ultimately also leads to a different ranking for the corresponding countries. In conclusion, a closer examination of the EECA region demonstrates that differences in the underlying democracy definition can and do lead to differing country ratings. In applications covering time periods and regions such as EECA prior to 1990 this can potentially be a source of results varying with the democracy measure used.

United States of America

A closer examination of the timeseries of United States will illustrate differences in temporal variation, in ranking and the disaggregation possibilities between the three indices.

Figure 3.8 displays the democracy index values and yearly rankings of the United States between 1972 and 2016. The US is coded as a perfect democracy receiving the highest possible value throughout the entire time series by both, Polity2 and FHI. This lack of variation for the two indices is also found in most highly democratic/autocratic countries.[105] In circumstances in which the research question at hand concerns such countries this feature needs to be kept in mind.

The differences in ranking between FHI and Polity2 are minor. However, the difference to Polyarchy is rather large. From 1999 to 2001 US Polyarchy drops by about 5.8%. What caused this sharp decrease? Plotting the components of Polyarchy for the United States over time (see Figure 3.9) shows that the clean elections index, v2xel_frefair,[106] is the main driving factor behind it.

[104] Since no disaggregate data is available for FHI and the checklist questions are vague at best, its coding of the EECA region cannot be discussed further here.

[105] A table displaying the within and between country variation for Polity2 can be found in the Appendix.

[106] "Question: To what extent are elections free and fair? Clarifications: Free and fair connotes an absence of registration fraud, systematic irregularities, government intimidation of the opposition, vote buying, and election violence.", Coppedge et al. (2017a), p. 58.

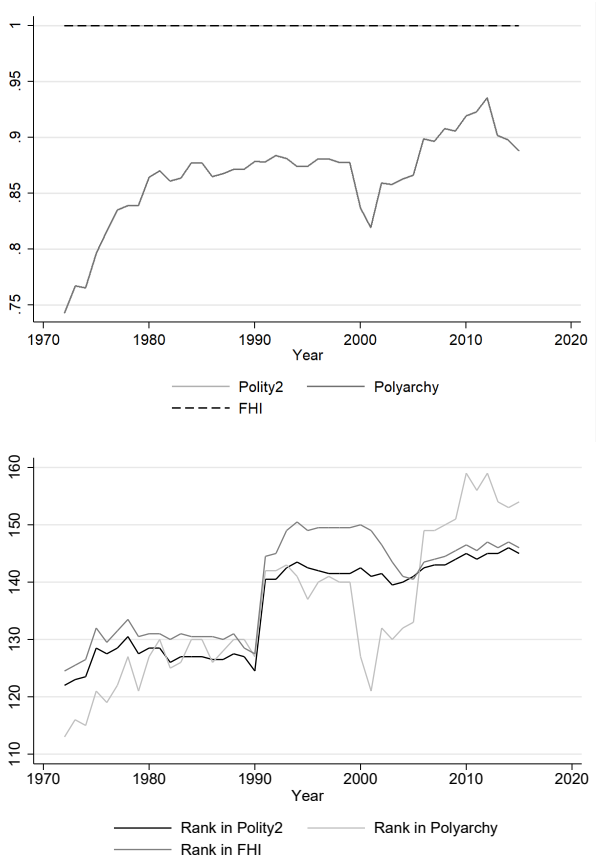


Figure 3.8: Democracy index values (left) and yearly rankings (right) for the United States of America from 1972-2016

Further decomposing the clean elections index into its subcomponents[107] (see Figure 3.10) exposes EMB autonomy (v2elembaut),[108] election voter registry (v2elrgstry),[109] election vote buying (v2elvot-buy),[110] election other voting irregularities (v2elirreg),[111] and election free and fair (v2elfrfair)[112] as driving factors (see top graph).

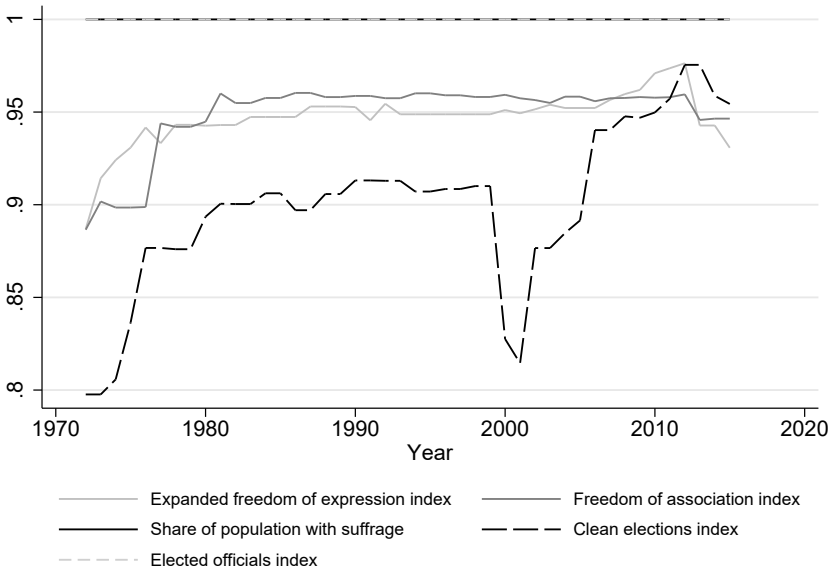


Figure 3.9: Components of Polyarchy for the United States over time

In sum, the drop in Polyarchy (and subsequently in the ranking of the US in comparison to other countries in the year 2000) is due to the 2000 presidential election and is not captured by Polity2 or FHI. This difference in the US- values and ranking for each of the indices illustrates how Polyarchy is more sensitive than Polity2 or FHI, which is both, an advantage (higher variability, more information is included) as well as (slight) flaw of the index. As discussed in section *Varieties of Democracy (V-Dem) Dataset* V-Dem does not provide a thorough discussion of the de facto and de jure elements included in the democracy measure. De facto (as opposed to de jure) indicators by nature rely more on judgement calls and inherently suffer from an increased bias. The indicators driving the decrease in US Polyarchy values address the de facto aspect of the United States democracy in 2000. US Polyarchy drops from around 0.88 in 1999 to 0.84 in 2000. Whether this sharp decrease is an appropriate reflection of changes in US levels of democracy around the 2000 election or whether it might be influenced/amplified by ideological judgements inherent in de facto indicators can be debated. It comes down to the question "what share of the underlying democracy definition consists of de facto vs. de jure attributes of

[107] Note that most of these indicators only occur in election years and are then repeated over election regime periods.

[108] "Question: Does the Election Management Body (EMB) have autonomy from government to apply election laws and administrative rules impartially in national elections?", Coppedge et al. (2017a), p. 86.

[109] "Question: In this national election, was there a reasonably accurate voter registry in place and was it used?", Coppedge et al. (2017a), p. 90.

[110] "Question: In this national election, was there evidence of vote and/or turnout buying?", Coppedge et al. (2017a), p. 94.

[111] "Question: In this national election, was there evidence of other *intentional* irregularities by incumbent and/or opposition parties, and/or vote fraud?", Coppedge et al. (2017a), p. 95.

[112] "Question: Taking all aspects of the pre-election period, election day, and the postelection process into account, would you consider this national election to be free and fair?", Coppedge et al. (2017a), p. 103.

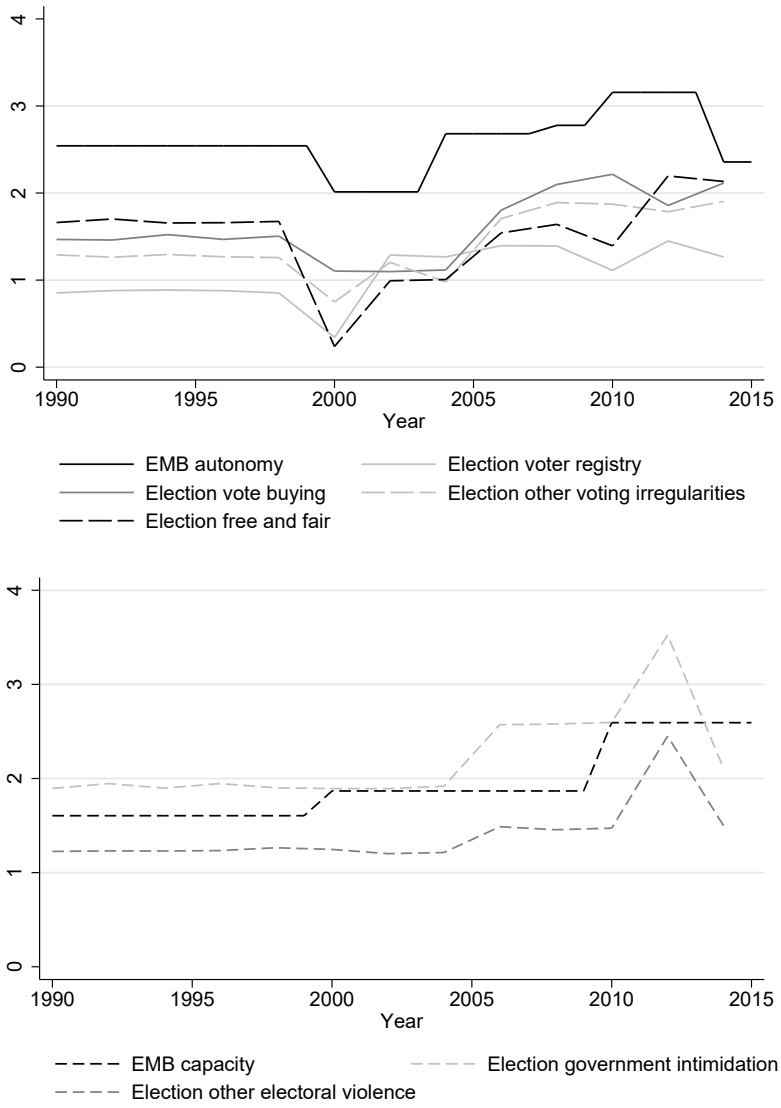


Figure 3.10: Subcomponents of the clean elections index for the United States of America from 1990 - 2016

a polity?". None of the three democracy measures properly address this question.

Nevertheless, Polyarchy still outperforms the other two measures with respect to all points discussed in the paper. Polyarchy's increased sensitivity/variation (especially for highly democratic/ autocratic) countries can be a large advantage in a time-series cross-country setting over the countries which are coded as 1/0 for decades by Polity2 or FH. The one thing the V-Dem project could improve is to find a system of "checks and balances" for de facto attributes so as to not let this sensitivity get out of hand.

3.4 Discussion

The three democracy indices studied in this paper were introduced at different times and for different purposes. In section Definition of democracy the distinction between *de facto* and *de*

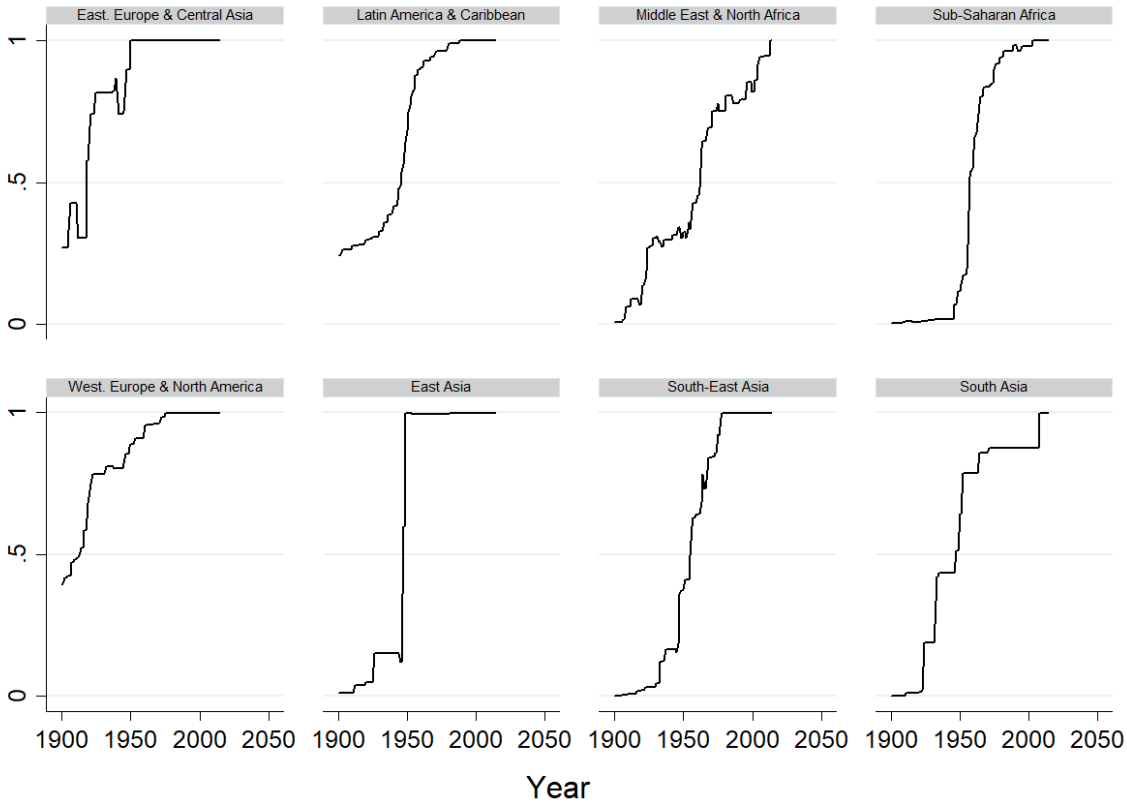


Figure 3.11: Average regional shares of population with suffrage, 1900 - 2016. Source: V-Dem indicator *v2x_suffr*

jure attributes included in a democracy measure was discussed. It ties into the very definition of the underlying concept of democracy/authority patterns. The Polity data was created to examine the durability of institutional frameworks. Its aim was to capture the *de jure* framework. FHI on the other hand was originally designed as a measure of civil liberties and political rights. It therefore is is more of a *de facto* measure. V-Dem's Polyarchy is situated somewhat in between the two hypothetical *de facto* and *de jure* endpoints. There is no "optimal" partition of *de facto* and *de jure* elements to incorporate in a democracy measure. The research question of the project at hand and the corresponding perception of democracy should determine the levels of *de facto* and *de jure* attributes. None of the three democracy indices are "better" or "worse" because of their choice of partition. While the choice in itself does not affects the indices overall validity it does affect the values coded. The country studies provided in the previous section show that the inclusion of more or less of such *de facto* and *de jure* elements leads to different country assessments by the three measures. The example of the coding of the United States highlights how the inclusion of *de facto* attributes can be a main driver behind different ratings across indices. Empirical researchers planning to work with democracy measures hence need to contemplate what degree of *de facto* and *de jure* attributes is appropriate for their research.

Throughout the paper the loss of variation in particular for countries at the very top or bottom of the scales has been discussed. The two country studies in the previous section highlighted this for Polity2 and FHI in the United States as well as for the suffrage dimension in Eastern Europe & Central Asia. The latter example raises the question whether suffrage today is even an important indicator to include in a measure of democracy as there is little/no geographical

variation. Figure 3.11 displays average shares of population with suffrage by region. The graphs show little or no variation in the last decade of the time series but a lot of variation in the first half of the 20th century. Thus the question posed above can be amplified in scope by asking: how should democracy measures adapt to the changing importance of certain dimensions of democracy over time? The share of population with suffrage has been an important dimension of democracy. It helped distinguish between different kinds of regimes and to capture some extent of the people's options to participate in the political process. Today, any new polity coming into existence will find it difficult to allocate suffrage to white adult males only, for example. The share of the population with suffrage is not as helpful in distinguishing between different polities today as it was 50 years ago. It goes well beyond the scope of this paper to provide a solution to the changing importance of certain democracy dimensions over time. Nevertheless, it is within the scope of this paper to point out that V-Dem excels at providing very precise and clear cut distinctions between a large number of dimensions. By providing the disaggregate data it also gives the empirical researcher the opportunity to include or disregard dimensions as needed.

3.5 Conclusion

This paper compares the three most commonly used democracy datasets: the PolityIV, the Freedom House and the V-Dem dataset. In a first step, it analyzes their respective democracy measures' underlying definition, their measurement scale, their data collection as well as the theoretical justification of their aggregation procedure to assess their validity and reliability. In addition, important strengths and weaknesses of each measure are discussed. The democracy indices from the V-Dem dataset surpass the other two democracy measures in all areas mentioned above. In a second step, this article compares the indices' respective coding of those observations included in all three datasets. There is a relatively high level of agreement in the country coding as well as in the country ranking between the indices for these cases. In conclusion, the most substantial differences between the indices lie in their coverage, the availability of disaggregate data and the above mentioned key areas. These are the central aspects for scholars to consider when choosing a democracy measure for their research.

Last but not least, the massive efforts undertaken and the equally enormous contributions of each of the three data projects discussed shall be underlined. As stated in previous sections each data project originated with quite different objectives and at different times. The first V-Dem dataset was released about 40 years after the first Polity and Freedom House data came out. In addition, it had been prepared by scholars familiar with the Polity and Freedom House data who knew the caveats and pitfalls of these datsets by heart. The V-Dem project was able to build their data on the foundations of scientific discourses about empirical democracy measurement since the introduction of Polity and Freedom House data. By offering extensive documentation, disaggregate data and sensitivity to new information through the Bayesian item response models the V-Dem data is optimally equipped to become the new standard in democracy measurement and to adapt to future challenges.

Chapter 4

'Tis but thy name that is my enemy: on the construction of macro panel datasets in conflict and peace economics

The empirical analysis of datasets covering a large number of countries and time periods has become an integral part of conflict and peace economics. As such, numerous studies examine relationships between and among macroeconomic, political, and conflict variables and this often involves the merging of disparate datasets to combine relevant variables for which the country unit of analysis, however, is not necessarily the same. This article highlights difficulties in the data merging process and, by way of example, presents detailed country coding unit comparison for two economic (UN Comtrade and World Development Indicators), two democracy (Polity IV and V-Dem), and two conflict datasets (UCDP/PRIO Armed Conflict Dataset and COW Militarized Interstate Disputes Dataset). We find that merging datasets can result in the elimination of very large numbers of observations due to unmergeable records and that dropped observations often include the very countries or territorial entities most of interest in conflict and peace economics.

JEL codes: F140, N400

Keywords: Comparative Economic History, International Economic History, Multicountry Studies, Comtrade, World Development Indicators, Varieties of Democracy (V-Dem), PolityIV, UCDP/PRIO Armed Conflict, Militarized Interstate Disputes

This chapter is published as: *'Tis but thy name that is my enemy: on the construction of macro panel data in peace economics.* Boese, V. A. and K. Kamin. Economics of Peace and Security Journal, 2019: 14(1), 5-26.
DOI: http://dx.doi.org/10.15355/epsj.14.1.5

Acknowledgements:

This chapter is joint work with Katrin Kamin, Christian-Albrechts-Universität zu Kiel. The authors would like to thank Scott Gates and Philipp Großkurth for excellent comments and suggestions.

4.1 What's in a name?

In conflict and peace economics, the construction of large panel datasets nowadays forms the basis for the majority of empirical cross-country studies. Originating from different sources, such panel datasets contain measures on variables such as international trade, economic growth, GDP, armed conflict, democratization, and government effectiveness.[1] But bringing these variables together, that is merging them into a single dataset, hinges on the exact identification of the country unit under study. To permit reasonable statistical inference, the country unit for which, for example, the trade value is calculated, should respond to the same entity for which all other variables in the dataset are coded. Unfortunately, the names, and even the physical borders, with which countries are coded vary considerably across different data sources.[2]

At the core of the coding differences lies the question "What's in a (country) name"? We argue that there are two complementary parts to the answer. The first regards the entity under observation, the unit of analysis: What is a country? The answer depends on the research framework. For example, the purpose of the Russett et al. (1968) state list as well as of the original Gleditsch and Ward (1999) state list was to capture recognized states in the international system. This particular definition of a country is of utmost relevance in analyses of authority structures. Nevertheless, one cannot blindly assume that the unit of analysis, that is, the country, is defined along the same criteria in economic or political datasets. Unfortunately, the burden of comparing the unit of analysis underlying different macro panel datasets lies with the scholar(s) attempting to merge them. As a consequence, we emphasize the importance of discussing the merging process in empirical studies in conflict and peace economics. The second part to the "What's in a (country) name?" question concerns the entity's label: Numerous scholars have presented ways to adjust for differences in country labels. For example, Hensel (2001) provides a thorough list of alternative historical state names and Heather Ba has created Stata files allowing for the mapping of country names, Correlates of War (COW) codes, and World Bank codes.[3]

That inconsistent country names across different data sources pose a problem is widely known among scholars working with macro panel datasets. Major attempts to standardize worldwide country coding already were undertaken half a century ago by Russett et al. (1968) and almost twenty years ago by Gleditsch and Ward (1999). Nevertheless, several problems remain unresolved and, unfortunately—with the emergence of readily available software packages and codes—a discussion of "what is the (country) unit of analysis" has become almost unfashionable. In spite of its tediousness and complexity, the country merging process is generally not discussed in academic papers (or in their supplementary materials).

The contribution of this article is hence twofold: First and foremost, it shows that in spite of all country coding scheme standardization efforts and relevant software packages or codes, the problem of inconsistent country coding in macro panel datasets persist. We therefore want to re-raise awareness of this problem and encourage a discussion of it in empirical cross-country studies in conflict and peace economics. Second, by way of illustration, in the Appendix to this article we provide overview tables of some of the gravest discrepancies in country coding across datasets which facilitate quick cross-dataset comparisons of country units.

[1] Examples of studies using such merged datasets include Hegre (2001), Blomberg and Hess (2006), Gates et al. (2006), Martin et al. (2008), Glick and Taylor (2010), Acemoglu et al. (2019), Dunne and Tian (2015), and d'Agostino et al. (2018).

[2] Hence the title of this article. 'Tis but they name that is my enemy (Act II, Scene ii, Shakespeare (2003)). Note: although the authors do not feel the same passion towards merging panel datasets as Juliett felt towards merging with Romeo, they do feel strongly about inconsistent country coding and the possible loss of information associated with it and hope that their dataset merging efforts will culminate in a different ending.

[3] See `http://heatherba.web.unc.edu/data-code/`.

4.1.1 Typology of inconsistencies

A typology of inconsistencies Inconsistent country names are the tip of the merging iceberg. Not only do names differ, but so does for example the period of existence for some countries. And worse, the documentation on the country coding schemes provided by the data projects is often sparse and contains errors.[4]

The following three types of inconsistencies between country units in different data sources and coding schemes are frequently observed and examined in this article:

1. A state name exists in one dataset but not in the other

   (i) Reason: different years (time series do not match and some states do not exist anymore/yet).
   - Example: when merging PolityIV with Comtrade data the Orange Free State cannot be merged as it ceases to exist before coding of Comtrade Data starts)
   - Result: country is unmergeable and drops out because it does not exist in one dataset

   (ii) Reason: different definition of statehood
   - Example: some datasets do not code Palestine as they do not consider it to meet formal requirements of statehood
   - Result: country is unmergeable and drops out because it does not exist in one dataset

   (iii) Reason: different state names (labels) or entities/territories (see inconsistency 3 described below)
   - Example: Yugoslavia and its successors are coded vastly different in terms of names and years across all datasets. How should these countries/observations be aggregated to make them comparable across datasets and to not loose conflict observations?
   - Result: country may drop out if no action is taken

2. A country is coded under the same name, but for different years in two datasets (time series for given country are not identical in both datasets)

   (i) Reason: missing observations within time series
   - Example: In V-Dem Germany, 1946-48, is not coded since the institutional framework of Germany during those years does not meet the formal criteria for the definition of their democracy indices
   - Solution: depends on application and on underlying assumptions made about reason for missingness, possibly interpolation

   (ii) Reason: country starts or ceases to exist and first/last year is not coded consistently across datasets
   - Example: PolityIV codes former East Germany between 1945-1990 whereas V-Dem codes it from 1949-1990.
   - Solution: depends on application, possibly extrapolation

3. A country is coded under different names:
   3A: for the same years in two datasets
   3B: for different years in two datasets

[4]For discussion, see the sections on democracy, economic, and conflict data in this article.

(i) Reason: it is clearly the same state only the label is different. This is often the case for 3A - or for 3B in combination with inconsistency type 2, reason (ii).

- Example: "St." vs. "Saint" or official vs. colloquial state names ("Plurinational State of Bolivia" and "Bolivia")
- Solution: use relevant Stata and R packages for renaming

(ii) Reason: the different names might refer to different underlying entities/ territories

- Example: We provide detailed overviews of these cases in Table C.3 (Democracy Datasets) and Table C.7 (Economic Datasets) of the Appendix.
- Solution: The option 3B case is by far the most difficult case as the years do not provide additional evidence on the actual entity captured. The question of how these entities could be compared in a meaningful way across datasets has no straightforward answer, rather the answer is case dependent.

Inconsistent country coding of types 1 to 3 lead to missing values in the final, merged dataset.[5] In this article we show that the extent of these "missing values" (they are not really missing, just missing due to inconsistencies) is vast and of particular relevance to empirical research in conflict and peace economics. Most country coding schemes differ in the naming and dating of a specific set of countries: Countries which have experienced armed conflict are less democratic and less trade open than the consistently coded ones. As a result, a merged dataset can contain a comparatively high share of missing values for this set of countries. Thus, it can no longer be considered a random sample. To minimize "missings," and to avoid losing valuable information, the process of creating large panel datasets should therefore be done with utmost care.

In general, there are three approaches to code countries in macro panel data: By (string) country names, by numeric code, or by alphabetic code. The most common schemes include (but are not limited to) the COW country list, the Gleditsch/Ward state list, and the ISO 3166 list of country codes.[6] In theory, numeric and alphabetic codes should facilitate the merging process. Unfortunately, several numeric and alphabetic codes schemes exist and often they are neither implemented consistently nor are the country codes easily translatable to each other. In R the package "countrycode" and in Stata the package "kountry" help with these issues.[7] These packages map country names and codes from one kind of macro country codes to another. They come with a slight disadvantage, though, as "[t]he mapping between the available dataset_names [types of country coding schemes] is not always perfect."[8] This is especially dire when using a comparatively new dataset such as V-Dem which does not follow any of the coded country schemes exactly. In addition, this assumes that each source dataset correctly applies the country coding scheme it is based on. In the following sections we show that this is not

[5]Note the difference between missing values and missing observations. For example, on the one hand, in the V-Dem dataset version 8 there are no observations for Germany between 1945 and 1948, leaving the panel unbalanced. In the World Development Indicators, on the other hand, the panel provided is balanced, that is, there is one observation for each country in each year. However, for a number of years the variable of interest contains a missing value. Ultimately, when merging two such sources and using the final dataset for statistical analysis, missing values and missing observations come down to the same thing: missing information. For most regressions or other analyses, software like Stata disregards observations whenever they contain missing values.

[6]COW: A country coding scheme employed by several of the macro panel datasets studied in this article. Data can be obtained from `http://www.correlatesofwar.org/data-sets/cowcountry-codes`. There are three variables: numeric and alphabetic country codes and statename. The dataset covers 217 countries. The country list includes 26 duplicate observations. Gleditsch/Ward: The Gleditsch and Ward (1999) state list builds on and revises the COW country list. First published in 1999, a current version is available at `http://ksgleditsch.com/statelist.html`. ISO: See `https://www.iso.org/iso-3166-country-codes.html`.

[7]R: See `https://cran.r-project.org/web/packages/countrycode/countrycode.pdf`. Stata: See Raciborski (2008).

[8]Quote:Raciborski (2008) (p. 392). Raciborski (2008) continues with a short overview of the most striking inconsistencies.

the case for several datasets. By letting Stata or R packages adjust the country names, the renaming—and subsequently the merging process—is put into a black box, inherently making it more vulnerable to mistakes.

We aim to take this data merging process out of its black box and use actual country names to prevent merging mistakes. In what follows we provide a detailed comparison of six datasets covering the indicators trade, democracy, and conflict. For each dataset a table with actual country names and years in the data is provided (see Boese and Kamin, 2018a, 2018b). These tables present an overview of the gravest discrepancies in country coding and allow for quick cross-dataset comparisons of country units. In addition, this article gives an overview of the extent of the country coding problem by comparing structural properties of the set of inconsistently coded countries to those of the uniformly coded ones and by discussing missing data as well as differences in annual coding.

On the one hand, this article provides assistance to scholars merging several source datasets. On the other, it highlights naming inconsistencies between data documentation, such as code books, and actual observations in the data. Such inconsistencies potentially lead to merging problems when blindly using the Stata or R packages (and the country coding scheme specified in the documentation) discussed above. We have the highest respect for all the data projects discussed in this article. We therefore hope that the lists of these inconsistencies are also of assistance to the data projects in aligning their documentation to their respective datasets.

The following three sections respectively provide thorough comparisons of two democracy, two trade, and two conflict datasets, including detailed tables comparing the country coding units. The article closes with a discussion of the results.

4.2 Democracy Data

This section compares the country coding units of two democracy datasets: V-Dem version 8 and the PolityIV dataset 2016. The tables referenced in this section can be found in the Appendix as well as in Boese and Kamin (2018a). We first discuss the countries listed in V-Dem version 8, then discuss the countries in the PolityIV dataset 2016, and then compare characteristics of the observations listed in both datasets with those listed in only one of the datasets.

4.2.1 V-Dem Data Version 8

The V-Dem dataset used for this article is V-Dem data version 8, in country year format. The variable of interest is the Electoral Democracy Index, v2x_polyarchy. V-Dem identifies the countries either by name, alphabetical country id, or numerical country id.[9] These country identifiers do not correspond to any of the prevailing country schemes implemented in the Stata or R packages mentioned above. To facilitate the merging process, we therefore provide a detailed list of county coding units in the data[10] and compare it to the country list in the V-Dem code book (Coppedge et al. (2018a) p. 36).

V-Dem excels in terms of transparency and provides a supplementary article on "V-Dem Country Coding Units v8" which lists and discusses all polities and countries and the respective years for which they are coded as well as a detailed explanation of the country borders used in the

[9] Alphabetical country_text_id: "Abbreviated country names". Numerical country_id: "Unique country ID designated for each country. A list of countries and their corresponding IDs used in the V-Dem dataset can be found in the country table in the codebook, as well as in the V-Dem Country Coding Units document".

[10] See Boese and Kamin (2018a), worksheet "V-Dem Codebook vs. Data".

coding.[11] It also provides detailed information on years in which a country is not coded (with the variables gapstart and gapend). However, there are several observations for which v2x_polyarchy is missing. Worksheet "Overview" in Boese and Kamin (2018a) shows the number of years for which each country is coded in V-Dem version 8, as well as its gaps (by coding decision) and its additional missing values.

For ten countries the names in dataset and documentation do not match.[12] These name mismatches are by no means a purely alphabetical problem. Take, for example, Vietnam. While there is no country named Vietnam, North or South, in the V-Dem dataset there is a "Republic of Vietnam" (coded from 1802–1975) and a "Democratic Republic of Vietnam" (coded from 1945–2017). The V-Dem Country Coding Units document, however, provides a detailed overview of the polities forming part of:

"Vietnam, South (35) Coded: 1802–1975. History: (...) Republic of Vietnam (also known as South Vietnam) (1955–1975)." and "Vietnam, North (34) Coded: 1945– History: Democratic Republic of Vietnam (i.e. North Vietnam) [declared] (1945); Democratic Republic of Vietnam (1945–1949); Democratic Republic of Vietnam [independent state] (1949–). Note: From 1976, the polity also includes areas formerly belonging to Republic of Vietnam (South Vietnam)."[13]

Take another example. In the documentation the numerical country id (365) is coded for two countries: Oldenburg, 1789–1867, and Saxe-Weimar-Eisenach, 1809–1867. In the dataset, however, only Saxe-Weimar-Eisenach is assigned country_id 365 while Oldenburg is assigned code 364.

4.2.2 Polity IV

A second dataset, capturing political authority patterns worldwide and over long periods of time, is the PolityIV project's dataset on "Political Regime Characteristics and Transitions, 1800–2016" (for short, the PolityIV dataset).[14] In the dataset countries are identified by their name, an alphabetic country code, or a numeric code.[15] These identifiers supposedly follow the COW country coding scheme.[16] Table 4.1 displays the results from merging the PolityIV data with the COW country list, finding that 13 percent of the countries are unmergeable when merging by country name, 6 percent when merging by numeric code, and 10 percent when merging by alphabetic code.[17] The unmergeable groups largely consist of countries of particular interest in conflict and peace economics such as the Koreas, Congos, Germanies, and Serbias. As a consequence, when merging the PolityIV data using a software package taking the dataset to be

[11] See Coppedge et al. (2018b).

[12] They are: Democratic Republic of Congo, Democratic Republic of Vietnam, German Democratic Republic, Mecklenburg Schwerin, North Korea, Republic of Vietnam, Republic of the Congo, South Korea, São Tomé and Príncipe, and Timor-Leste.

[13] V-Dem Country Coding Units, p. 27.

[14] See Marshall et al. (2017b).

[15] Alphabetic: The variable scode ("Alpha Country Code: Each country in the Polity IV dataset is defined by a three-letter alpha code, derived from the Correlates of War's listing of members of the interstate system" (Marshall et al. (2017a), p. 12). Numeric: ccode (numerical, "Numeric Country Code: Each country in the Polity IV dataset is defined by a three-digit numeric code, derived from the Correlates of War's listing of members of the interstate system" (Marshall et al. (2017a), p. 11).

[16] Supposedly: See Marshall et al. (2017a) (p. 11).

[17] To be clear, the share of unmergeable countries is calculated as: number of unmergeable countries/ total number of countries in PolityIV (i.e., $26/195 \approx 13.3\%$, $11/194 \approx 5.7\%$, and $19/194 \approx 9.8\%$. Note that the rows are labeled correctly although one could in fact omit "and COW" from the second row since, if countries are mergeable in a merge between COW and PolityIV, they must exist in both datasets. In the first row, however, are unmergeable countries only, i.e., those which exist only in the PolityIV dataset.

Merging by	Country name	Numeric code	Alphabetic code
Unmergeable countries in PolityIV	26	11	19
Mergeable countries in PolityIV & COW	169	183	177

Table 4.1: Number of (un)mergeable countries in a merge of the PolityIV Dataset with the COW country list

Dataset	A: V-Dem	B: PolityIV
total number of observations	26,537	17,228
total number of nonmissing observations	24,115	16,992
number of countries	201	195
number of years	1789-2017	1800-2016

Table 4.2: Description of democracy datasets

in "COW coding scheme" these countries may not be properly dealt with.

It is worth noting that country names and alphabetic and numeric codes are not coded consistently over time within the PolityIV dataset, i.e., there are 195 different country names, but only 194 different alphabetic and numeric codes. This is not due to a single country having different names and only one code, but to a number of countries and several code/label constellations. Examples include Yugoslavia (either ccode 345 and scode YUG or ccode 347 and scode YGS; that 347 and YGS also are used for Serbia and Montenegro in the dataset further complicates matters), Ethiopia (either ccode 529 and scode ETI or ccode 530 and scode ETH), Pakistan (either ccode 769 and scode PKS or ccode 770 and scode PAK). Further, ccode 860 and scode ETM is used for East Timor and Timor Leste, and ccode 255 and scode GMY is used for Germany and Prussia. Additionally, in the PolityIV dataset we note duplicate observations for Yugoslavia in 1991 and for Ethiopia in 1993. This further complicates the merging process as the scholar is forced to decide how to proceed with the duplicates.

4.2.3 Comparison of the democracy data

Table 4.2 describes both democracy datasets. The variable of interest in each dataset is a democracy index: v2x_polyarchy for the V-Dem data and polity2 for the PolityIV data.[18] The total number of nonmissing observations refers to the number of observations for which the respective variable of interest contains nonmissing values.

When merging the datasets by country name and year, observations of inconsistency types 1 to 3 cannot be merged. Table 4.3 shows the number of mergeable and unmergeable observations by source dataset. As discussed, even though an observation might be listed, the variable of interest can contain a missing value. Hence the lower half of Table 4.3 proves the same information for all observations with nonmissing values. To make the number of observations comparable across datasets in Table 4.3, only observations from the time period covered by both datasets are considered (that is, V-Dem observations before 1800 as well as the year 2017 were left out to match the PolityIV time series). Around 41 percent of the V-Dem and around 9 percent of the PolityIV observations cannot be merged. To assess whether the unmergeable observations are systematically different from the mergeable ones we calculated the average levels of democracy for each group. Table 4.4 shows the results of two t-tests, one for V-Dem, one for PolityIV.

[18] V-Dem's v2x_polyarchy: Range 0 to 1 (most democratic). PolityIV's polity2: Range –10 to +10 (most democratic).

Merging	A: V-Dem	B: PolityIV
unmergeable observations only in A	10,929	-
unmergeable observations only in B	-	1,619
mergable observations in both:	15,609	
non-missing observations only in A	9,380	-
non-missing observations only in B	-	1,571
non-missing mergeable observations in both:	14,736	15,421

Table 4.3: Merging V-Dem and PolityIV data

Dataset:	A: V-Dem		B: PolityIV	
average level of democracy unmergeable group	0.1377		-0.4495	
average level of democracy mergeable group	0.3428		-1.5493	
difference in average democracy levels between groups	0.2051	***	1.0998	***

Table 4.4: Two sample ttests of average level of democracy.
Democracy variable V-Dem: v2x_polyarchy, range: 0 (most autocratic) to 1 (most democratic)
Democracy variable PolityIV: polity2, range: -10 (most autocratic) to 10 (most democratic)
Note: *** Statistically significant at the 1% level.

In both datasets, the unmergeable group had a significantly lower average level of democracy (To be clear, the t-tests were carried out only on the nonmissing observations noted in Table 4.3).

4.3 Economic data

UN Comtrade[19] and the World Bank's World Development Indicators (WDI)[20] contain economic data. We first discuss the countries listed in the UN Comtrade data, then those in the WDI, and then compare the country coding schemes of both datasets. The tables and worksheets referenced to in this section can be found in the Appendix as well as in Boese and Kamin (2018b).

4.3.1 UN Comtrade

The indicator taken from UN Comtrade is total exports in current U.S. dollars from each country to the rest of the world. The Comtrade dataset is an unbalanced panel as it only contains years for which countries have reported trade. Hence, time series differ from country to country. The first year for which some countries reported trade is 1962, the last year is 2017 (few observations are available for the start and end years of the time series). Comtrade offers data coded according to two different systems for international trade statistics: The Harmonized System (HS), introduced in 1988, and the Standard International Trade Classification (SITC), introduced in 1962, with the latter being less detailed than the former. To obtain the longest possible time series, we concatenated SITC classification export data, 1962–1987, with HS classification export data, 1988–2017.

In addition to gaps in the time series caused by missing observations (as discussed above) the export variable contains missing values for several observations. Missing information primarily

[19] See "United Nations Comtrade Database" (2018).
[20] See The World Bank (2017b).

indicates that trade was not reported and is not to be equated with zero trade flows.[21] This is crucial concerning the tackling of zero trade flows and appropriate model choice.[22]

The country name abbreviations of the official UN country list[23] correspond to the country names used in the Comtrade data with the exception of Côte d'Ivoire and Réunion, which contain spelling errors in the downloaded Comtrade dataset ("C√¥te d'Ivoire" and "R√©union").

4.3.2 World Development Indicators

The economic indicator taken from the World Bank's WDI is trade openness, defined as the percentage share of trade of each country's GDP, that is, (imports+exports)/GDP. Starting in 1960, the time series runs to 2016. The distinction between zero trade and missing data in the WDI is equivalent to the one in UN Comtrade. In contrast to Comtrade, however, the WDI data is a balanced panel with one observation for each country and year. Nevertheless, trade openness contains missing values for several observations due to missing information on GDP, exports, or imports. In addition to countries, WDI provides aggregated information on country groups (such as "Europe & Central Asia" or "Low & Middle Income"). These where taken out of the list to facilitate reading (the full list of country groups removed is available in Boese and Kamin, 2018b, worksheet "Disregarded Country Groups").
To our knowledge, the World Bank does not provide an explicit country coding scheme upon which WDI data are based. However, the World Bank does provide a list of countries upon which the World Integrated Trade Solution (WITS) data are based.[24] It is unclear whether this list also forms the basis of the WDI dataset. Of 15,048 observations in the WDI dataset used in this article, 30 percent (4,560 observations) do not match the WITS list. Several of them are due to naming inconsistencies such as, for example, "Bahamas, The" versus "Bahamas".

4.3.3 Comparing the economic data

In a comparison of the economic datasets[25] the sheer number of naming inconsistencies[26] and single appearances of countries (that is, they appear in one, but not in the other dataset)[27] stands out. Additional cases, difficult to handle when merging datasets, are countries that started and ceased to exist, yielding different country names for different or the same territories and for different years (inconsistency type 3). While WDI refers to each country under one name continuously for the entire time series, this is not the case for the UN Comtrade data. In Comtrade, countries are coded by different names and years. Table C.7[28] displays the cases where this kind of inconsistency is in place. The table shows that Comtrade distinguishes the underlying country entities in much more detail. There is, for example, only one "Germany" in the WDI data as opposed to "Germany", "Fmr Fed. Rep. of Germany" and "Fmr Dem. Rep. of Germany" in the UN Comtrade data.

[21] For a discussion of missings in trade data see, for example, Keshk et al. (2010) (Section 3.3, p. 10), Barbieri et al. (2009) (p. 476), and Boehmer et al. (2011).
[22] See, for example, Santos Silva and Tenreyro (2006).
[23] The UN provides a list of country codes and names at `https://unstats.un.org/unsd/tradekb/Knowledgebase/50377/Comtrade-Country-Code-and-Name`.
[24] https://wits.worldbank.org/wits/wits/witshelp/content/codes/country_codes.htm.
[25] See Boese and Kamin (2018b), worksheet "Overview".
[26] See Table C.4 or Boese and Kamin (2018b), worksheet "naming inconsistencies" for inconsistency type 3, reason (i) (one country coded with different names but for the same year and years).
[27] See Tables C.5 and C.6 or Boese and Kamin (2018b), worksheet "existence asymmetry" for inconsistency types 1 and 3.
[28] Also see Boese and Kamin (2018b), worksheet "inconsistency type 3".

Dataset	A: Comtrade	B: WDI
total number of observations	12,768	15,048
total number of nonmissing observations*	6,790	10,643
number of countries	228	264
number of years	1962-2017	1960-2016

Table 4.5: Description of trade datasets
Note: *The total number of nonmissing observations refers to the number of observations for which the respective variable of interest contains nonmissing values.

Merging	A: Comtrade	B: WDI
unmergeable* observations only in A	3,803	-
unmergeable observations only in B	-	6,083
mergable observations in both:	8,965	
non-missing observations only in A	1,449	-
non-missing observations only in B	-	3,765
non-missing mergeable observations in both	5,341	6,878

Table 4.6: Merging Comtrade and WDI data
Note: *When merging both datasets by country name and year those observations of inconsistency types 1-3 are unmergeable.

Assuming that the ending of one state and the beginning of a new one are coded in detail through the year variable by WDI, can the country coding units be supposed to be the same across the two datasets? The sparsity of country coding unit documentation renders it impossible to answer this question. There is no information on whether territories changed, and on whether or how much this change was incorporated in the coding. This becomes a severe drawback to the data when complementary variables for the analysis of trade flows, such as country size, GDP, measures of distance and—most importantly—borders are taken into account.[29]

The case of Sudan (see Table C.7)[30] illustrates the problem: WDI codes "South Sudan" and "Sudan". For the latter, the measure of trade openness is available for the whole time series (1960–2016). For "South Sudan", the indicator is available from 2008–2015. UN Comtrade codes "Sudan" (2012–2015) and "Former Sudan" (1963–2011, with gaps). Hence, WDI takes 2008 as the year of birth for "South Sudan", while Comtrade (implicitly, because it does not code "South Sudan" as a country)[31] codes a new state "Sudan" from 2012 onward. Similar cases are Serbia (with or without data for Kosovo or Montenegro) and China (with or without data for Hong Kong, Macao, and Taiwan).[32]

The country name by itself does not allow for an exact indication of the territory coded. In a statistical analysis only of trade, it might not matter whether Sudan or South Sudan is included. In conflict and peace economics, however, where relationships among conflict, politics, and economics are of high interest, such lack of accuracy effectively becomes an impediment to an appropriate econometric analysis.

[29] Anderson and Van Wincoop (2003), for example, demonstrated that national borders are a highly important impediment to trade.

[30] Boese and Kamin (2018b), worksheet "inconsistency 2.0", rows 36–38.

[31] The fact that no "South Sudan" is included in the UN Comtrade data is itself somewhat astonishing since trade data is available (otherwise WDI would not be able to code it).

[32] See World Bank (2017, p. XVII).

Dataset:	A: Comtrade	B: WDI
average level of trade variable unmergeable group	$2.72 \cdot 10^{14}$	66.16
average level of trade variable mergeable group	$3.98 \cdot 10^{13}$	76.14
difference in average total export & trade openness levels btwn both groups	$-2.32 \cdot 10^{14}$***	9.98***

Table 4.7: Two sample ttests of average level of trade and trade openness.
Trade variable Comtrade: total exports, TradeValueUS, range: US$ 37,310 to US$ $2.34 \cdot 10^{16}$
Trade variable WDI: trade openness, tradeop, range: 0 to 860,8 (in %)

Table 4.5 describes both trade datasets. For Comtrade, the variable of interest is total exports in current U.S. dollars (TradeValueUS); for the WDI data, it is trade openness as a percentage of GDP (tradeop). Table 4.6 shows the number of mergeable and unmergeable observations by source dataset. As discussed, even though an observation might be listed the variable of interest can contain a missing value. Hence the bottom half of Table 4.6 provides the same information for all observations with nonmissing values. To make the number of observations comparable across datasets in Table 4.6 only observations from the time period covered by both datasets are considered, i.e., 1962–2016. About 30 percent of the Comtrade observations, and about 40 percent of the WDI observations, cannot be merged.[33] To assess whether the unmergeable observations are systematically different from the mergeable ones, we calculated average levels of total exports and trade openness for each group. Table 4.7 shows the results of two sample t-tests: For Comtrade, the average export level is statistically significantly higher (given the exponent) in the unmergeable than in the mergeable group. For WDI, the unmergeable country group had a significantly lower level of average trade openness.

Looking at the naming inconsistencies (see Table C.4 of the Appendix) confirms this "higher-lower" difference: The high levels of export values in the unmergeable group in Table 4.7 are driven by observations from the U.S., Germany, Macao, and Hong Kong.[34] Table 4.7 hence provides a good intuition to the effects of inconsistent country coding: Either the cases of high export levels or of low trade openness are lost due to merging problems. Either one is problematic in terms of statistics and, depending on the analytic aim, might lead to biased estimates.

4.4 Conflict data

In theory, the datasets for economic and political variables code each variable for all years during which a country exists. The conflict datasets, however, are fundamentally different: By design, they only code conflict variables for years in which a conflict occurred in a given country and which surpassed some conflict criteria (for example, 25 battle-related deaths). Consequently, time series and cross-section data dimensions contain gaps for country-years without armed conflict.

The UCDP Armed Conflict dataset version 18.1 (Pettersson and Eck (2018); also see Gleditsch et al. (2002); UCDP, 2018) studies armed conflict above a yearly threshold of 25 battlerelated deaths. The Militarized Interstate Disputes (MID) B dataset version 4.2 (Palmer et al. (2015)) captures militarized interstate disputes which can involve, for example, a display of force without incurring any battle deaths. Therefore, the gaps in the datasets will be very different, and merging them by country and years coded does not provide insights on, or a comparison

[33] Again, to be clear: $3,803/(3,803+8,965) \approx 29.7\%$ and $6,083/(6,083+8,965) \approx 40.4\%$.

[34] This is shown in Boese and Kamin (2018b), worksheet "Unmergeable Outliers Comtrade". It contains all unmergeable Comtrade observations sorted by export values

Merging by	Numeric code	Alphabetic code
Unmergeable countries in MID B	4	4
Mergeable countries in MID B & COW	191	191

Table 4.8: Number of (un)mergeable countries in a merge of the MID B Dataset with the COW country list

of, country coding units. Nevertheless, both datasets acknowledge the importance of defining country coding units. In the remainder of this section, we show that even within each of these datasets there are inconsistencies between the country coding units as defined by the respective data project and the actual observations in the data. As a result, these observations are either dropped, potentially falsely matched, or have to be manually adjusted when using Stata or R commands for merging countries.

4.4.1 UCDP/PRIO Armed Conflict dataset version 18.1

The UCDP/PRIO Armed Conflict dataset acknowledges the importance of a precise description of country coding units[35] and dedicates an entire section of its code book[36] to the exact definition of country coding units. It includes a country table with numerical and alphabetical country codes, state names, and start and end years for the countries that form part of the international system of states. Tables C.8 to C.12 list the countries coded in the actual data and compares them to the system membership table from the UCDP/PRIO code book. The system membership table must include more observations since, by definition, it also includes countries without armed conflict. But Tables C.8 to C.12 show that even when restricted to countries with armed conflict there are inconsistencies in the country names (for example "Burkina Faso" and "Burkina Faso (Upper Volta)", "DR Congo (Zaire)" and "Congo, Democratic Republic of (Zaire)", and "Ivory Coast" and "Cote D'Ivoire").

4.4.2 MID B version 4.2

The MID B version 4.2 dataset includes one observation per participant to a militarized dispute, 1816 –2010, with countries taken from the Correlates of War (COW) list. The MID B dataset itself does not contain (string) state names. Instead, countries are coded with a three-digit numerical code (ccode) and with an alphabetical code (stabb). Before joining variables from the MID B dataset with any other macro panel data, such as WDI, a first step therefore is to merge MID B with COW, but four countries cannot in fact be merged (Table 4.8. The three-digit alphabetic codes for these countries are RUM, USR, VTM, and ZAI.

This is a perfect example of the difficulties associated with merging by country as it is hardly possible to determine with certainty which underlying entity (territory) is exactly covered, for example, by USR or VTM. This also illustrates why, for this article, we chose to employ merging by country (string) names, not codes. VTM could stand for (Democratic) Republic of Vietnam, Vietnam North, Vietnam South, or Vietnam. While the exact entity coded remains unclear, it is very clear that this case contains information relevant for studies of conflict. That the MID B dataset states that it follows the COW country list convention when in fact it does not, makes it effectively impossible to determine for some observations which actual underlying entity is considered a country during which period of time.

[35] "The definition of a state is crucial to the UCDP/PRIO conflict list" (UCDP/PRIO Armed Conflict Dataset Codebook, 2018, p.13).

[36] See Section 4: "System Membership Description" (UCDP/PRIO Armed Conflict Dataset Codebook, Themnér (2018), p. 13).

4.5 Discussion and conclusion

Large-scale cross-country datasets are frequently merged in quantitative studies in conflict and peace economics. We find that the coding of country units overlaps across datasets only for a relatively small proportion of countries. Discrepancies in country naming or other forms of country identification such as numerical or alphabetical country IDs are frequent among countries splitting up or (re)uniting during the time period studied. Examples include Yugoslavia, Germany, Vietnam, and Sudan. If the names are not adjusted, these inconsistencies render such observations unmergeable and, when joining variables from several data sources, ultimately result in missing values. When these missing values then are dropped from an analysis, important information is lost. This loss of information is of particular severity in conflict and peace economics as countries which split up or reunite often do so accompanied by armed conflict and thus contain valuable information.

The dataset comparisons made in this article demonstrate that inconsistencies in country coding across macro panel datasets remain a relevant challenge in cross-national studies. They show that for economic datasets as well as democracy datasets the unmergeable group is of a large size (up to about 40 percent of all observations) and significantly differs from the group of mergeable observations. In particular, the group of unmergeable countries is on average less democratic than the mergeable group. Depending on the economic measure analyzed (and, with it, the country naming scheme applied), a group of countries with high exports or another group of countries with low trade openness cannot be merged.

These discrepancies can be attributed, in part, to differences in country labels. Several projects, such as Hensel (2001) and the aforementioned software codes and packages can help adjust them. However, another part of the inconsistent country coding is due to different perceptions and definitions of the unit of analysis. The exercises carried out for this article show that the actual entity captured can differ by source dataset. While this makes creating merged panel datasets consisting of economic, political, or armed conflict factors challenging in its own right, proper merging might be a necessary condition for analysis. For an armed conflict dataset, relevant state units might differ significantly from datasets on democracy or trade flows (the coding of Palestine, Hong Kong, or Macao are examples). As a result, the burden of discussing the unit of analysis studied and of ensuring that countries correspond to the same entity across merged datasets, lies with the individual scholar or team. This article encourages scholars to discuss the merging process in their academic papers (or supplementary materials) and to not take the problem of inconsistent country names lightly. This is particularly the case in conflict and peace economics, where relevant information is systematically lost when unmergeable observations are discarded.

Furthermore, it is worth noting that country names are not the only dimension of macro panels to be carefully compared across datasets before merging. It goes well beyond the scope of this article to additionally compare the actual time periods covered. However, we point out that the time dimension underlying the calendar year coding of macro panels does not necessarily coincide with the actual calendar year. To quote from the World Bank: “In most economies the fiscal year is concurrent with the calendar year ... Most economies report their national accounts and balance of payments data using calendar years, but some use fiscal years” (The World Bank (2017a), p.117). Time inconsistencies, then, are another potential source of erroneous inference, in particular when studying the effect of conflict on the economy or the political system, or vice versa.

Last, but not least, we pay tribute to the creators of the datasets discussed in this article: Assembling and maintaining these datasets is a Herculean task. The challenges associated with

inconsistent country names and units across datasets can, however, lead to serious consequences in conflict and peace economics. Unfortunately, while an easy solution to the noted problems is not likely to exist, given the different purposes each of the source datasets is created for, we hope that our comments here increase broader awareness and discussion of these problems and that our tables in the Appendix (and online) facilitate quick cross-dataset comparisons of country coding.

Chapter 5

Heterogeneity matters: on the dynamic interactions of trade, development, democracy and conflict

The past decade has witnessed a return to protectionist measures as well as a global rise in nationalist movements. Understanding the economic and political effects of such changes in trade policies or levels of democracy has gained renewed importance. There is a vast amount of literature studying the bilateral relationships between international trade flows, democracy, development and conflict. While it finds strong evidence for correlations between the four factors so far no conclusive evidence has been presented as to the direction of causal links between the factors. This paper employs a country specific vector autoregressive model allowing for endogenous dynamic interactions between trade, democracy, development and conflict. More specifically, it analyzes how shocks in one of these variables affect the others over time to investigate "what causes what". The dataset used covers 68 countries and the years 1960 to 2016. Results confirm the presence of simultaneous effects from all variables on one another. In addition, effect size and sign is substantially heterogeneous across countries providing strong evidence against the validity of the homogeneous slope parameter assumption.

JEL codes: F14, F51, N40, O5, P45

Keywords: international trade, democracy, development, international conflict, panel vector autoregression, endogeneity

This chapter received the Michael D. Intrilligator Best PhD Student Paper Award at the 23rd International Conference on Economics and Security in Madrid, Spain (June 2019).

Acknowledgements:

This chapter is joint work with Katrin Kamin, Christian-Albrechts-Universität zu Kiel. The authors would like to thank Scott Gates, Steffen Elstner, Matthias Hartmann, Håvard Hegre and the participants of the brownbag seminar 2018 at the Department of Peace and Conflict Research, Uppsala University, for excellent comments and suggestions. Visualization of Figure 5.1 by Christian Isachsen. The impulse response functions were graphed with stata colorscheme *plotplainblind* by Bischof (2017).

5.1 Introduction

The study of each of the bilateral relationships between democracy, development, trade openness and armed conflict[1] constitutes and entire field of research. In each of these fields there is no consensus about the direction of the causal link. If there is one thing to be taken away from this literature it is that democracy, development, trade openness and armed conflict have shown to be endogenous to each other. This endogeneity should also be accounted for in statistical models. Until now the literature has usually used two types of models to examine the bilateral relationships: either (dynamic) panel data models, usually estimated with system or difference GMM approaches, but not accounting for the indirect effects these four variables have on each other, or simultaneous equation models which can take into account the indirect effects but do not contain autoregressive parameters. However, when it comes to factors like democracy, development, trade or conflict a country's state in the last period is of high importance in determining its current state.

At the same time, the past ten years have seen a global rise in nationalism and protectionism on the one hand, and an erosion of democratic norms in several countries on the other hand. With it an understanding of the role of economic interdependence and democracy for a peaceful interconnected world gains renewed importance: how does the international system respond to changes in trade or in governance systems?

This paper employs a vector autoregressive (VAR) model that allows to control for theses indirect effects. It models the dynamic relationships between democracy, development, trade and conflict in a globalized world. More specifically, it analyzes how shocks in one of these variables affect the others over time and examines the direction of effects for each bilateral relationship. Impulse response functions are used to examine the effects. These impulse response functions allow for contemporaneous correlation of errors and thereby control for endogenous effects between the four factors.

The results provide two major insights: firstly, they confirm that all variables affect each other. Therefore, indirect effects must be accounted for in any project examining the effect of one factor on another. Secondly, there is substantial heterogeneity in effect size and direction across countries. This suggests that the homogeneity of slope parameters assumption frequently employed in classic fixed effects models in cross-country analysis is inappropriate for capturing the effects of one of the variables on the others.

The rest of the paper proceeds as follows: Section 5.2 outlines the mechanisms relevant to the identification of effects in this paper. Variables and data used are introduced in Section 5.3. Section 5.4 states the hypothesis to be tested and describes model and identification strategy. Central findings are presented in Section 5.5. Model robustness is assessed in Section 5.6. The findings are discussed in Section 5.7 and a conclusion can be found in Section 5.7.

5.2 Why endogeneity matters

The interactions between democracy, development, trade and conflict are depicted graphically in Figure 5.1: the right-hand side displays what is best described as a "political cycle" evolving around democracy, while on the left-hand side the "economic cycle" evolves around trade. Both

[1] For readability and for brevity, this paper might refer to trade openness as "trade" and to armed conflict as "conflict". However, note that the underlying definitions of these terms correspond to how they have been defined in Section 5.3.

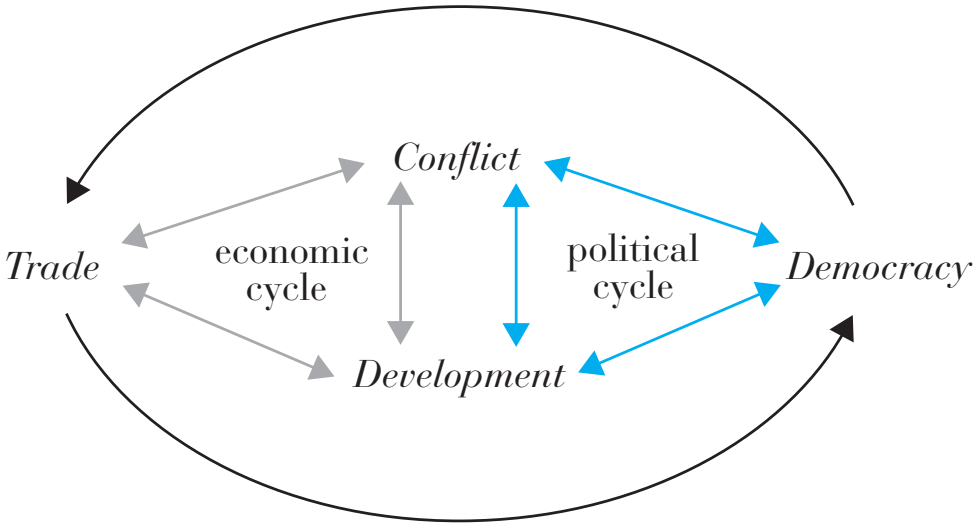


Figure 5.1: Mapping the endogenous interactions of trade, democracy, development and conflict

cycles show how closely the four factors are interrelated. Economic and political science literature have studied the endogeneity and causal bilateral relationships between each possible pair of the four factors. The most important findings are described in the following.

Trade and Development

Frankel and Romer (1999) point out the endogenous relationship between trade and development. Using income per capita, the authors find that trade raises income. Others have assessed the distributional consequences of trade: Helpman et al. (2010) find that aggregate inequality increases when countries are nearly symmetric due to reinforced within-sector effects, while Egger and Kreickemeier (2012) introduce intergroup-inequality and find that while aggregate welfare increases, so does inequality within as well as between the groups. The mechanism at work can be described as follows. Increasing trade makes a broader selection of goods and services accessible for the population which also leads to knowledge transfer, increased human capital and higher levels of development. In turn, human capital is a crucial factor for trade, both in terms of imports as well as exports: increased human capital on the labor market leads to a rise in production possibilities concerning export goods, while increasing income will boost spending capacity and demand for imported goods, Yanikkaya (2003).

Trade and Conflict

Conflicts affect trade through different channels: Resources are redirected towards defense expenditures (known as the guns vs. butter trade-off, (see e.g. Anderton and Carter (2009), Anderson and Marcouiller (2002)), resources and goods are destroyed and production possibilities shrink due to conflict, and future and present economic activities are disrupted (e.g. capital flight, increase in transport costs, etc.) which leads to a decrease of material well-being (see e.g. Long (2008)). Overall, opportunity costs are rising if trade gains are lost: utility and income decrease if consumption shifts away from preferred goods and production shifts away from areas of comparative advantage. Empirical studies have found mixed results depending on conflict data used. Trade as well affects conflict: according to the liberal theory trade promotes peace.

While the majority of empiric literature supports liberal theory (Blomberg and Hess (2006), Martin et al. (2008), Keshk et al. (2010)), realists argue that trade may also spark conflict (see e.g. Barbieri and Levy (1999), Barbieri et al. (2009)).

Conflict and Development

Greed or vertical economic inequality may spur conflict (see Collier and Hoeffler (2004)). Furthermore, in conflict infrastructure is destroyed and water and sanitation points deteriorate. Goods and services are not available in the same way as in times of peace. Living conditions deteriorate, accompanied by an increase in diseases and malnutrition. If such conditions persist, the gap between elites and population can widen and the society can wind up in a conflict trap (Collier (2003)).

Democracy and Development

Lipset (1959) describes democracy as a system that forms a political culture of negotiation and compromise. Within that framework, democracy acts as a system of redistribution of income but also of increased education. This stimulates human capital accumulation as well as labor productivity which in turn act as an engine for economic growth (see e.g. Acemoglu et al. (2014), Boucekkine et al. (2016), Rodriguez and Rodrik (2000)). A stable and growing economy with increasing income will itself lead to a more educated population demanding increased participation, redistribution of power, rights and income. As such, socio-economic development has been shown to be both outcome and determinant of democracy (see Acemoglu et al. (2019), Doucouliagos and Ulubaşoğlu (2008), Madsen et al. (2015), Narayan et al. (2011).)

Democracy and Conflict

The vast democratic peace literature (see e.g. Acemoglu and Robinson (2005), Oneal et al. (1996), Gates et al. (1996), Mitchell et al. (1999)) has shown that democracy is a system for the peaceful resolution of conflict. Hegre (2014) gives a good overview on this part of the literature. Conflicts in turn affect democracies with the impact being determined by the type conflict. Boese (2015) shows for example, that revolutionary conflict over the past 50 years has had a positive effect on democracy. Acemoglu and Robinson (2005) show that the distribution of power between elites and population play a critical role in the development of a state towards democracy or dictatorship.

Trade and Democracy

The liberal theory suggests that democracies trade more (Mansfield and Snyder (2002)): On the one hand, elites have less opportunities to extract protectionist rents, as democracies tend to support rather the preferences of the greater part of the consumers than those of a few producers (Tavares and Wacziarg (2001)). Free trade in turn consolidates democracy. It reduces protectionist rents and thereby reduces the incentives for authoritarian groups to seek power. With increasing democracy, both exporter and importer signal compliance with the rule of law, reliability in business processes and a higher product quality (Liu and Ornelas (2014) and Yu (2010)). However, groups that benefit from protectionism often try to impair politics via lobbying (Tavares and Wacziarg (2001)).

Why democracy, development, trade and conflict?

The dynamic interactions between trade, development, democracy and conflict form central pillars of the literature outlined above. In addition, Subramanian and Satyanath (2004) find that trade openness, conflict and democracy are strongly (positively) correlated with macroeconomic

stability. Rodrik et al. (2004) model the interactions between trade openness, democracy, geography and income to assess the relative impact of the three former variables in determining income levels. The theoretical logic underlying their paper is very close to the one applied here. The model used in our paper considers the same endogenous variables and (due to the substantial effects of conflict found in the democratic and liberal peace literature) adds conflict as an endogenous variable.[2]

Endogeneity

Given the amount of findings on any of these bilateral relationships two points suggest themselves: first, the presence of simultaneous bilateral effects and second (through said simultaneous direct effects) the presence of indirect effects. Both, in consequence lead to interaction structures such as those in Figure 5.1. Several scholars have noted both, indirect effects, e.g. Persson and Tabellini (2009), Tavares and Wacziarg (2001) Rodrik et al. (2004), Baum and Lake (2003), and simultaneous effects, e.g. Persson and Tabellini (2009), Rodrik et al. (2004), Russett and Oneal (2001). However, this article is the first systematic study that acknowledges both points and that allows for a) all four factors being jointly determined, b) the factors to simultaneously affect each other and c) indirect effects in their econometric specification.

[2]Since geography is a time-invariant characteristic it is captured in each country's intercept in our model.

The effect of...	...on: Conflict	Development	Democracy	Trade
Conflict	-	Mac Ginty and Williams (2016) Justino et al. (2013) Besley and Persson (2010)	Hegre (2014) Boese (2015) Russett and Oneal (2001)	Blomberg and Hess (2006) Martin et al. (2008) Kamin (n.d.)
Development	Collier and Hoeffler (2004) Collier (2003) Gartzke and Weisiger (2014)	-	Lipset (1959) Acemoglu et al. (2008) Rigobon and Rodrik (2005)	Yanikkaya (2003) Keesing (1967)
Democracy	Collier and Rohner (2008) Crescenzi and Kadera (2016) Hegre (2014)	Acemoglu et al. (2019) Boucekkine et al. (2016) Pozuelo et al. (2016)	-	Mansfield and Snyder (2002) Liu and Ornelas (2014) Yu (2010)
Trade	Barbieri et al. (2009) Hegre et al. (2010) Martin et al. (2008)	Frankel and Romer (1999) Rigobon and Rodrik (2005) Rodriguez and Rodrik (2000)	Rigobon and Rodrik (2005) Adsera and Boix (2002)	-

Table 5.1: Selected studies on the relationships between democracy, development, trade and conflict

5.3 Data and descriptive statistics

In this paper V-Dem's Liberal Democracy Index, *v2x_libdem*, is used as a measure of democracy.[3] Bernhard et al. (2017) show that the operationalization of democracy affects the results. A major drawback of the vast amount of empirical literature carried out on the interrelationships between democracy and other variables is that several studies use similar measure(s) of democracy (see Altman et al. (2018), p. 14), most notably the Polity2 and Freedom House Index. Boese (2019) provides a comprehensive introduction into quantitative democracy measurement and a detailed overview of the empirical analyses these two indices should not be employed for. She concludes that the comparatively new democracy indices by the Varieties of Democracy Institute (V-Dem)[4] were created to answer to most of the problems posed by "classic" democracy measures, such as Polity2 and Freedom House Index. Consequently, they substantially outperform them in terms of measure validity and reliability (this includes, for example, their underlying definition of democracy, their measurement scales or the theoretical justification of their respective aggregation procedures). Therefore, V-Dem's Liberal Democracy Index is employed in this paper. Its democracy definition includes the dimensions participation, contestation and constraints on the executive decision making authority.[5]

This paper uses the armed conflict categorization provided by the Uppsala Conflict Data Program (UCDP). A conflict is coded as such once a threshold of 25 battle-related deaths is reached. Armed conflict has four conflict sub-categories: Extra-systemic armed conflict, interstate armed conflict, internal armed conflict and internationalized internal armed conflict.[6] Furthermore, the armed conflict category provides information on conflict involvement of different sides to a conflict. To include all conflict observations a binary variable was created that is equal to one once a country is involved in any type of conflict. Thus, all conflict involvements - no matter on which side of the conflict a country stands - are included.

Against the background of the country-level of analysis, economic interdependence is operationalized as trade openness. Trade openness has been widely used as a measure of economic integration (see e.g. Rodriguez and Rodrik (2000)). Trade as percent of GDP is taken from the World Bank Indicators, who define the indicator as follows: "Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product." Data is available as weighted average and on annual level.

Literature including variables on socio-economic development often uses GDP per capita. Since this study includes trade openness as measure for economic interdependence and hence as an economic measure, we abstract away from including a "socio-economic" quantification, but rather introduce a measure for development. We quantify development by using the World Development Indicator "Life expectancy at birth, female". The indicator is defined as "the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life." Data is available as weighted average and on annual level. Mortality rates mirror the health conditions of a country and are therefore a commonly used indicator for development.

[3] *v2x_libdem* is taken from the V-Dem Dataset Version 7.1, Coppedge et al. (2017d).

[4] The V-Dem indices are available since Lindberg et al. (2014).

[5] For a detailed introduction into democracy measurement as well as explanations of the concepts of participation, contestation and constraints see for example Boese (2019).

[6] The exact UCDP definitions of conflict types are provided in the Appendix

Variable	Mean	Std. Dev.	Min	Max
Liberal Democracy Index	40.84	29.76	1.22	90.34
Trade Openness	65.63	50.42	4.92	441.60
Life Expectancy, Female	66.49	12.58	29.28	87.14
Armed Conflict	0.30	0.46	0.00	1.00

Table 5.2: Summary statistics for pooled estimation sample

Variable	Std. Dev.	Min country means	Max country means
Liberal Democracy Index	26.46	4.41	87.62
Trade Openness	46.32	19.14	330.71
Life Expectancy, Female	11.13	45.12	80.52
Armed Conflict	0.24	0.02	0.89

Table 5.3: Variation between country means

Variable	Std. Dev.	Min	Max
Liberal Democracy Index	13.98	-56.67	37.12
Trade Openness	20.70	-101.52	192.82
Life Expectancy, Female	6.00	-22.72	20.03
Armed Conflict	0.39	-0.89	0.98

Table 5.4: Within country variation. The minimum (maximum) column refers to the lowest (highest) deviation a country exhibited from its respective panel mean.

5.3.1 Descriptive statistics

The dataset used in the analysis covers 68 countries and the time perid 1960 to 2016, i.e. 3,876 observations (57 observations per country). The dataset is balanced.[7] Table 5.2 provides the mean, standard deviation, minimum and maximum of the variables in the estimation sample.

Instead of taking the pooled sample as a basis, Table 5.3 examines the panel means of each country's time series. The variation between country means is quite substantial, especially for trade openness: the highest country mean is around 330 units whereas the country with the lowest mean displayed average trade openness levels of around 19. Since the armed conflict variable is a dummy equal to one in conflict years the minimum (maximum) value in Table 5.3 shows the minimum (maximum) percent of years in the time series a country was involved in armed conflict. The country with the lowest number of years involved in conflict was involved 2% of the 57 years, i.e. one year. The country with the highest number of years in armed conflict was involved during 89% of its time series (equivalent to 51 years).

How did the country values vary within the panels over time? Table 5.4 shows that the within variation for each country is substantially lower than the variation between the country averages for three of four variables (democracy, trade and development). In other words, the countries in this sample differ from each other (in terms of their trade, democracy or development levels) than each country's values vary over their respective time series. This heterogeneity of countries is remarkable and will be captured by the model used in this paper.

[7] A detailed list of countries and the respective regions included in the sample can be found in the Appendix (see Table D.1).

5.4 Model and Estimation

Findings of previous studies provide a starting point for this study: they suggest the presence of simultaneity, i.e. that trade, democracy, development and conflict each affect one another bilaterally. For each bilateral relationship, studies exist showing that "A causes B" as well as "B causes A". Note, the four factors bilaterally affect each other both directly as well as indirectly through the other variables. The model used in this paper allows trade, democracy, development and conflict to be jointly determined and captures the indirect effects. As such it provides a suitable econometric framework to test the following hypothesis:

Hypothesis: Development, democracy, trade and conflict simultaneously affect each other

Econometric implications of this hypothesis are the following: First, to avoid omitted variable bias a model needs to include each of these factors. Second, imagine four separate "one equation"-models of each of these variables with regressors consisting of lags and current values of the other variables. In each of these equations neither the simultaneity nor the indirect effects that the variables have on each other through a third or fourth variable are accounted for. The errors are contemporaneously correlated across the four equations. By examining orthogonalized shocks, this model allows for such a contemporaneous correlation of error terms. Examining these shocks will show whether there are simultaneous effects from each variable on all others (once we account for indirect effects) or whether there are bilateral relationships for which the effects solely run "from A to B" (and not vice versa). Is there a chronological order in which these factors affect each other? We assume that development and democracy are slow changing variables whereas trade and conflict react to shocks much faster in comparison.

Two limitations imposed by the dataset are worth noting. First, the 68 countries used in this study constitute but a part of the global system of states. For reasons of data availability it is not possible to include further countries. Data availability, however, is correlated with some of the four variables, most prominently conflict. Afghanistan, for example, could not be included as it's trade openness time series exhibited missing values for 25 of the 57 years used in the analysis. Nevertheless, the estimation sample consists of 68 countries from all regions of the world and as such is "as representative as possible". The second limitation is measure reliability. Socio-economic development is a latent factors difficult to measure. As such it is often measured in vastly differing ways. While we cannot adjust data availability, we do control for different specifications of development in the robustness checks section. Our main results are robust to these changes.

5.4.1 Reduced form VAR

The model used in this paper assumes that development, democracy, trade and conflict are jointly determined. To test whether there are simultaneous effects between each other the following reduced form VAR(4) model is estimated for each country i:

$$y_{i,t} = A_{i,1}y_{i,t-1} + ... + A_{i,4}y_{i,t-4} + \kappa_i + \epsilon_{i,t}, \tag{5.1}$$

$$i \in \{1, 2, ..., 68\}, t \in \{1, 2, ..., 57\}$$

where $y_{i,t} = (development_{i,t}, democracy_{i,t}, trade_{i,t}, conflict_{i,t})'$ is a $4x1$ vector of dependent variables, $A_{i,1}, ..., A_{i,4}$ are $4x4$ matrices of lag coefficients to be estimated and $\epsilon_{i,t}$ is a $4x1$ white noise process. To keep the model parsimonious the $4x1$ vector κ_i is the sole exogenous regressor (this permits a nonzero mean $E[y_i, t]$). With this model specification all current period measures of democracy, development, conflict and trade are a function of past values of each other. Hence,

the total reduced-form effect that a past increase in, for example, democracy had on each of the other dependent variables can be calculated and orthogonalized effects can be examined over time.

The number of lagged values to include in the equation deserves some attention. The data used in this paper is in annual format. Including one lag hence is equivalent to adding an entire year of information to the model (in light of the slow changing nature of variables such as democracy or development this is a fair amount of time). To make sure that even a more volatile measure such as trade openness is appropriately captured in the VAR model a lag length of 4 years is selected.[8] From a theoretical perspective, four years is an appropriate lag length choice as it is the average electoral period length in the countries under study. Standard Information Criteria (AIC, adjusted R^2) presented in Section 5.6 confirm this choice of lag length.

The following three assumptions one the error term are central for this model:

$E[\epsilon_{i,t}] = 0$, $E[\epsilon_{i,t}\epsilon'_{i,t}] = \Sigma$ and $E[\epsilon_{i,t}\epsilon'_{i,s}] = 0$, for $t \neq s$

Through the non-singular matrix Σ the error terms are allowed to be contemporaneously correlated. However, they must be uncorrelated with their lags and lead values as well as all right-hand side variables of the model.

Given the high amount of variation across countries (rather than within each countries time series)[9] the VAR(4) model is estimated country by country. This permits the slope parameters to be heterogenous across countries. In most cross-country analyses using panel data slope parameters are assumed to be homogenous across countries and systematic differences between countries are assumed to be captured through country specific fixed effects. Through estimating our model country-by-country it provides the slope parameters with the possibility to differ (i.e. the coefficients have the option to differ as much as to be homogenous across countries) while including a country specific intercept, κ_i. In that sense the model is an extension of above mentioned panel data models.

Using impulse response functions the dynamic relationships between democracy, development, economic interdependence and conflict can be depicted graphically. To derive them, it is useful to transform model (5.1) in its Vector Moving Average, VMA(∞), Representation:

$$\begin{aligned} y_{i,t} &= A_{i,1}y_{i,t-1} + ... + A_{i,4}y_{i,t-4} + \kappa_i + \epsilon_{i,t} \\ &= A_{i,1}Ly_{i,t} + ... + A_{i,4}L^4y_{i,t} + \kappa_i + \epsilon_{i,t} \\ (I - A_{i,1}L - ... - A_{i,4}L^4)y_{i,t} &= \kappa_i + \epsilon_{i,t} \end{aligned}$$

Let the lag polinomial $\Phi_{i,4}(L) := I - A_{i,1}L - ... - A_{i,4}L^4$ then

$$y_{i,t} = \Phi_{i,4}(L)^{-1}\kappa_i + \Phi_{i,4}(L)^{-1}\epsilon_{i,t} \tag{5.2}$$

For $\Phi_{i,4}(L)^{-1}$ to exist, $|\Phi_{i,4}(L)| \neq 0$. $\Phi_{i,4}(L)^{-1}\kappa_i = \mu_i$ where $E[y_{i,t}] = \mu_i$ (there is no t-subscript as the process is stationary)

$$y_{i,t} = \mu_i + \sum_{s=0}^{\infty} \theta_s \epsilon_{i,t-s} \tag{5.3}$$

[8] Robustness checks with up to ten lags can be found in the Appendix. The results do not change.
[9] See Section 5.3.1.

where $\theta_0 = I$ ans θ_s is a function of $\Phi_{i,4}(L)$. This Vector Moving Average, VMA(∞), Representation is the basis upon which impulse response functions are created. The effect of the shocks $\epsilon_{i,t}$ on $Y_{i,t}$ are now captured by

$$\frac{\delta y_{i,t}}{\delta \epsilon_{i,t-s}} = \theta_{s_i} \tag{5.4}$$

and can be interpreted as the effects of a shock in one variable on the others over time. Note that $y_{i,t}$ and $\epsilon_{i,t-s}$ are $4x1$ vectors. This allows us to calculate the response of the $n-th$ element in y to a shock in the $m-th$ element.

The impulse response functions depict the following situation: in the first period there is a one period (in this framework: one-year) positive shock with a magnitude of one standard deviation in the impulse variable. All other/response variables are set to their mean values in the first period. The impulse response graphs (see for example Figure 5.3 or Section D.3 of the Appendix) show the response variable's deviation from its mean in each year after the positive shock of the impulse variable in the first period. The solid line represents the point estimate of the response. The dashed lines highlight a 95 percent confidence band around that estimate. This paper focuses on responses to shocks over a ten year period.

5.4.2 Structural VAR

Given that the error terms are contemporaneously correlated in model (5.1) the estimated parameter values therein are impossible to interpret and further identifying assumptions are necessary. These assumptions in our case are that development and democracy are (in comparison to trade and conflict) slow changing variables, i.e. a shock in development or democracy will have an effect on trade and conflict in the same period while democracy and development will take time in responding to a shock in trade or conflict. Therefore, we assume the following ordering: development, democracy, conflict and trade.[10] That means development affects all other factors contemporaneously, but is not contemporaneously affected by shocks in the other variables. Democracy affects conflict and trade contemporaneously and is affected only by contemporaneous shocks of development and so on.

From an econometric perspective implementing these assumptions means model 5.1 needs to be orthogonalized (to obtain a variance-covariance matrix of error terms with orthogonalized zero off-diagonal elements). This paper uses a Choletsky decomposition, i.e.:

$$Cy_{i,t} = CA_{i,1}y_{i,t-1} + ... + CA_{i,4}y_{i,t-4} + C\kappa_i + C\epsilon_{i,t}, \tag{5.5}$$

where $C\epsilon_{i,t} = u_{i,t}$ and $Var[u_{i,t}] = \Sigma_u = \sigma^2_{u_i} I$.

This structual VAR(4) model is then transformed into its VMA(∞) Representations and the impulse response functions are calculated as discussed above. The next section summarizes the main findings from these orthogonalized impulse response functions.

5.5 Results

The hypothesis to examine is that simultaneous effects from variable A to variable B as well as from B to A are present. As there are four endogenous variables, there are 6 bilateral relationships to examine: 1. trade and democracy, 2. trade and development, 3. trade and conflict, 4.

[10] As Robustness Checks results for a different orderings are reported in Section 5.6.

development and democracy, 5. development and conflict and 6. democracy and conflict. Each bilateral relationship consists of two parts: A affects B and B affects A. If the impulse response functions display significant effects for both of these parts simultaneous effects are present and the hypothesis can be accepted.

The VAR(4) model is estimated and impulse response functions are calculated for each country, i.e. 68 times. There is one impulse response function for every country (68) and every bilateral relationship part ($6 \cdot 2 = 12$), i.e. per model specification $68 \cdot 12 = 816$ impulse response functions are calculated. The question of how to aggregate the information contained in these functions to a displayable level is not trivial. To test our hypothesis, for each part of the bilateral relationship, the number of countries exhibiting a significant effect is of immediate interest. Therefore, Figure 5.2 displays the number of countries with significantly positive and negative (and no) results for each impulse-response combination. Each cell of the table constitutes one part of a bilateral relationship.[11]

Figure 5.2 provides two fundamental insights: First, it suggests the hypothesis is true - simultaneous effects between all variables exist. For all combinations of variables "A affects B" as well as "B affects A". This is crucial. Any model estimating the effect of one of these variables on another has to account for the simultaneous relationships as well as the indirect effects. Second, for all impulse-response combinations there are positive as well as negative responses, i.e. the effects that these variables have on one another are far from homogeneous across countries. For each part of the bilateral relationships the responses differ across countries in terms of effect sign, magnitude and timing. Section D.3 of the Appendix contains the impulse response functions (as well as a discussion thereof) for countries with significant responses ordered by bilateral relationship. In fact, the responses are not even homogeneous within regions. Section D.2 of the Appendix shows that aggregating the impulse response functions on even a regional level disguises the variations contained within the country responses. A similar effect would occur had this model been estimated under a homogeneous slope parameter assumption.

[11] Note, that the question "how does a variable respond to shocks in itself?" is not part of this analysis. The corresponding impulse response combinations are reported in Figure 5.2 solely for the sake of completeness.

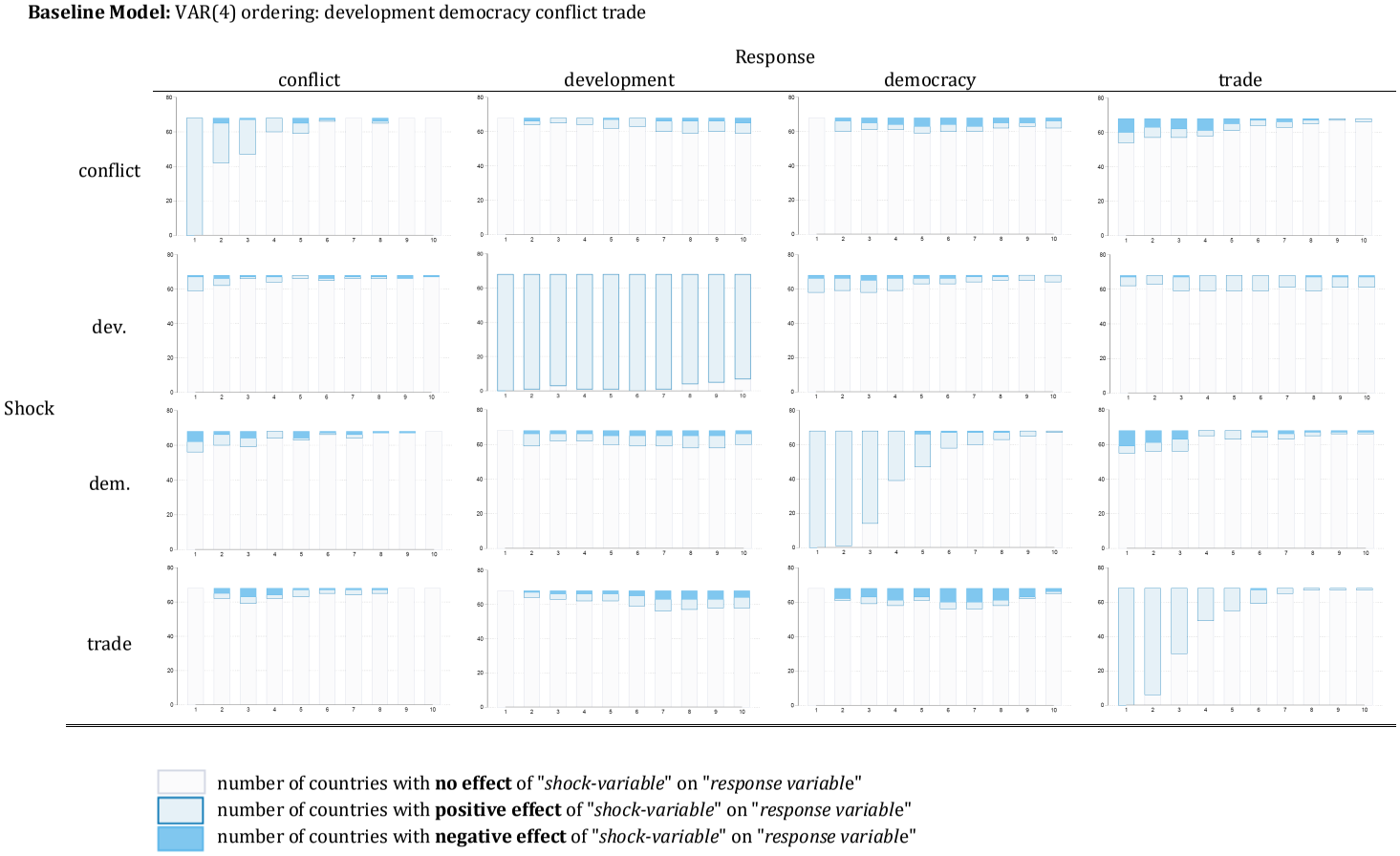


Figure 5.2: Number of countries with non/significant responses to the given shocks

The sole part of a bilateral relationship for which there is an overwhelmingly positive effect is that of development on trade openness. There are three countries for which a one period one standard deviation positive shock in development leads to negative responses: the Central African Republic, Uruguay and Nicaragua (see Figure 8). The substantial magnitude of the negative effect suggests that cross-country heterogeneity in responses cannot be neglected for this bilateral relationship either.

As depicted in Figure 3, even within one country the effect of a shock can be heterogeneous over time. In Kenya, for example, a one period one standard deviation positive shock in development leads to a significant positive effect for trade openness within the first three periods but switches to a significant negative and lasting effect in period six. For Portugal, a shock in development has the reverse effect: within the first three periods, trade is affected negatively, but switches to a positive effect in period six. These switches can be observed across regions and in the bilateral relationships between conflict and trade, development and democracy, and conflict and democracy.

5.6 Robustness Checks

How sensitive are these results to the identifying assumption and the model specification? To answer this question the model was re-estimated with six different specifications all of which are discussed in the remainder of this section.

In any VAR model the choice of lag length deserves attention. The baseline model of this paper was run with a lag length of four years since this approximately equals the average electoral period. To determine whether this lag selection is appropriate, several avenues were pursued: first, standard autocorrelation tests, such as the LM-test (Null hypothesis: no serial correlation of order 1-4) were consulted. As to be expected in a cross country setting such as this one autocorrelation is present in some of the countries. Eyeballing the residual plots for these countries (provided in Section D.5) suggest no systematic presence of heavy autocorrelation. Nevertheless, to make sure autocorrelation is negligible the VAR(4) model was re-estimated in first differences. Although the number of countries displaying significant results is smaller (see Figure D.26 of the Appendix) the same simultaneous effects as well as country specific heterogeneity in effect magnitude, timing and sign are visible. Comparing standard goodness-of-fit criteria for the baseline VAR(4) model in levels (Table5.7) and in first difference (Table 5.5) confirms the VAR(4) specification in levels. In the presence of autocorrelation an alternative option is to increase lag length. Therefore, the model was re-estimated as a VAR(8) and VAR(10) model. The responses remain heterogeneous across countries and the effects still appear simultaneously for both parts of each bilateral relationship (see Figure D.30 for VAR(8) and Figure D.31 for VAR(10)). Comparing model fit across these three lag specifications (see Table 5.6) suggests that longer lag specification yield better fits. However, with increasing lag length four countries must be dropped from the analysis because of colinearity in the variables. To include the highest number of countries possible and since the results remain robust the lag length of four remains the preferred specification.

By choosing female life expectancy as a measure of development this paper deviates slightly from common approaches to measuring development, namely by GDP-related variables. As robustness checks the model was re-estimated using once GDP per capita growth and once logged GDP per capita as measures of development. Before discussing the results it it worh remembering (see Section 5.3) that GDP is included in the measure of trade openness used in this paper:

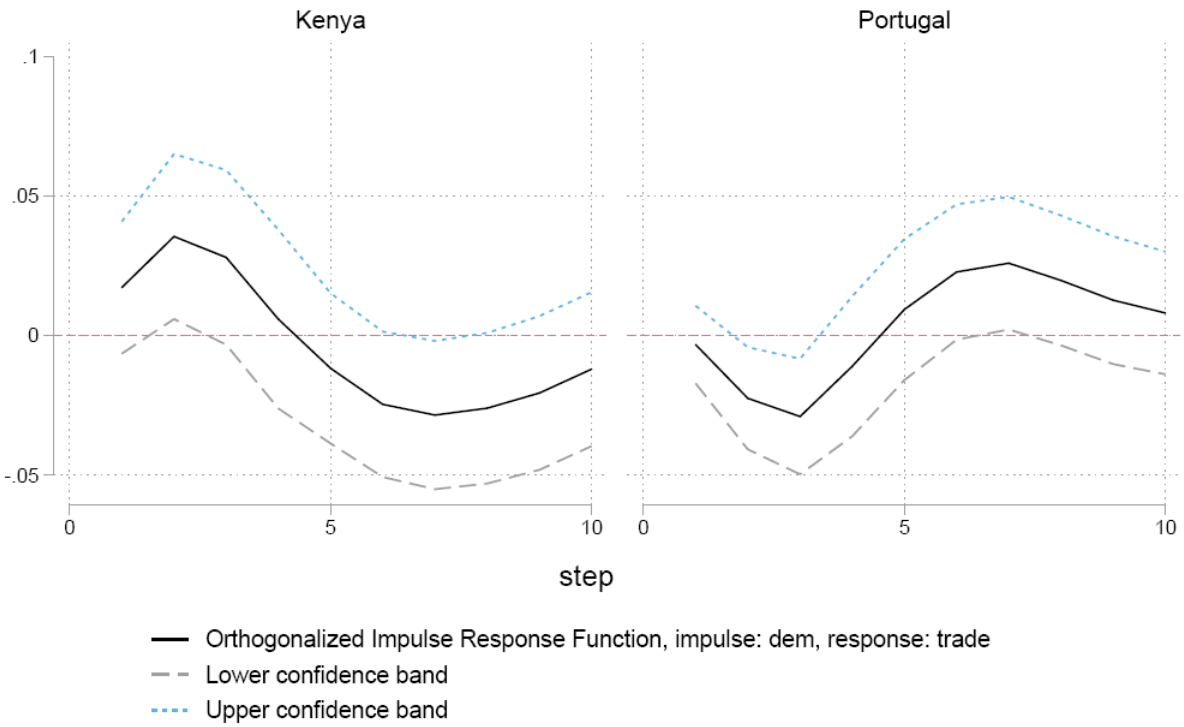


Figure 5.3: Impulse response functions for Kenya and Portugal displaying a switch from a significant positive effect of democracy on trade to a significant negative effect and vice versa.

$$trade_t = \frac{(exports_t + imports_t)}{GDP_t}$$

This implies that

$$\frac{\delta trade_t}{\delta GDP_t} = -\frac{(exports_t + imports_t)}{GDP_t^2} < 0, \qquad \text{if } exports_t + imports_t > 0 \text{ and } GDP_t \neq 0$$

as well as

$$\frac{\delta GDP_t}{\delta trade_t} = -\frac{(exports_t + imports_t)}{trade_t^2} < 0, \qquad \text{if } exports_t + imports_t > 0 \text{ and } trade_t \neq 0$$

GDP per capita growth and logged GDP per capita are defined as

$$growth_t = \frac{GDPpc_t - GDPpc_{t-1}}{GDPpc_{t-1}} \qquad \text{and} \qquad lnGPDpc_t = ln(GDPpc_t)$$

Assuming that $GDPpc$ is a function positively increasing in GDP the shocks examined in this paper would (ceteris paribus) and by pure definition of variables be expected to have the following effects:
A positive shock in development, measured by $growth_t$ or $lnGDPpc_t$, would correspond to an increase in GDP per capita (and consequently in GDP) leading to a decrease in $trade_t$. Conversely a shock in trade openness implies a reduction in GDP and with it a reduction in $lnGDPpc_t$ and $growth_t$, respectively. Interestingly, for both specifications (using $growth_t$ or $lnGDPpc_t$) positive responses to shocks in trade as well as to shocks in development are clearly visible. The number of countries with trade values responding negatively to shocks in development increases substantially in comparison to the VAR(4) model, which was to be expected given the definition of the variables. The number of countries exibiting signifcant results is smaller when the GDP

model in first differences			
Dep. Var.	Obs.	adj. R^2	AIC
development	68.00	0.71	-4.92
		(0.43)	(3.71)
democracy	68.00	0.11	4.51
		(0.23)	(1.71)
conflict	68.00	0.20	0.40
		(0.22)	(0.78)
trade	68.00	0.07	-1.64
		(0.17)	(0.88)
Overall			-1.86
			(3.19)

Table 5.5: Model fit for model in first differences. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported below each cell.

	Baseline VAR(4)			Lag length 8			Lag length 10		
Dep. Var.	Obs.	adj. R^2	AIC	Obs.	adjusted R^2	AIC	Obs.	adjusted R^2	AIC
development	68	1.00	-5.02	64	0.13	-5.26	64	1.00	-6.27
		(0.01)	(3.75)		(0.19)	(3.91)		(0.01)	(3.94)
democracy	68	0.90	4.34	64	0.89	4.05	64	0.92	3.03
		(0.11)	(1.69)		(0.11)	(1.79)		(0.10)	(1.93)
conflict	68	0.45	0.22	64	0.39	0.09	64	0.58	-0.90
		(0.29)	(0.74)		(0.35)	(0.86)		(0.37)	(1.48)
trade	68	0.83	-1.79	64	0.82	-1.96	64	0.87	-2.80
		(0.13)	(0.85)		(0.14)	(0.83)		(0.15)	(0.94)
Overall AIC			-2.48			-3.69			-9.09
			(3.23)			(3.73)			(4.63)

Table 5.6: Comparing model fit across different lag specifications. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported below each cell.

per capita growth specification is employed, which is also unsurprising since it is a growth rather than a levels variable (the response of development to a shock in trade, for example is closer to the baseline VAR(4) result for the $lnGDPpc_t$ specification than for the $growth_t$ specification). According to the information criteria presented in Table 5.7 across different specifications of development baseline VAR(4) remains the preferred specification.

The ordering of the variables in the Choletsky decomposition represents a central identifying assumption of the structural VAR. To assess the model's sensitivity to this assumption a Choletsky-decomposition was employed and impulse response function were calculated using the alternative ordering: democracy, development, trade and conflict.[12] The results (see Figure D.27) remain very similar.

[12]The two slow changing variables were switched and so were the two fast changing variables. Deviating from the slow- and fast-changing setup (i.e. ordering the variables with trade (or conflict) as first or second variable in the VAR) would imply that the a slow changing variable like democracy or development responded to a shock in trade (or conflict) within that same period, which seems unreasonable.

	Baseline VAR(4)			**GDP p.c. growth**			**log GDP p.c.**		
Dep. Var.	Obs.	adj. R^2	AIC	Obs.	adj. R^2	AIC	Obs.	adj. R^2	AIC
development	68	1.00	-5.02	67	0.13	8.02	68	0.97	-1.35
		(0.01)	(3.75)		(0.19)	(0.76)		(0.05)	(0.75)
democracy	68	0.90	4.34	67	0.89	4.41	68	0.90	4.35
		(0.11)	(1.69)		(0.11)	(1.73)		(0.10)	(1.70)
conflict	68	0.45	0.22	67	0.39	0.29	68	0.40	0.25
		(0.29)	(0.74)		(0.35)	(0.75)		(0.34)	(0.74)
trade	68	0.83	-1.79	67	0.82	-1.74	68	0.83	-1.77
		(0.13)	(0.85)		(0.14)	(0.84)		(0.15)	(0.84)
Overall AIC			-2.48			10.72			1.20
			(3.23)			(2.75)			(2.71)

Table 5.7: Comparing model fit across different specifications of development. "Obs." refers to the number of countries in the analysis. Values reported for adjusted R^2 and AIC by equation are pooled averages. Standard deviations are reported in parentheses below each cell.

5.7 Discussion & Conclusion

This paper estimates the simultaneous effects of trade openness, democracy, armed conflict and development on each other by employing a VAR model. The previous section confirms that the results are robust to a number of different model specifications and that standard information criteria confirm the model choice. The two central findings obtained are: i. the prevalence of simultaneous effects between development, democracy, trade and conflict and ii. the substantial heterogeneity in responses across and even within countries. These findings partially explain as to why the literature so far displayed findings in favor of each direction: first, because the effects are mixed and impacts run both ways of each bilateral relationship. Second, the contemporaneous determination of the four factors needs to be accounted for econometrically. Third, heterogeneity of responses across countries and over time needs to be allowed for.

A large part of the literature has accounted for simultaneity in the bilateral study of each of the four factors by, e.g. the use of lags, fixed effects, simultaneous equation models or IV approaches (see e.g. Rigobon and Rodrik (2005) and Martin et al. (2008)). However, the present study proves that solely examining bilateral relationships without controlling for time-varying other influences is not enough (as, e.g. seen in Frankel and Romer (1999) and Acemoglu et al. (2019). Simultaneous effects of more than two factors have to be accounted for.

Furthermore, the results of this paper strongly advocate against the widely assumed homogeneity of slope parameters. Section D.3 of the Appendix thoroughly discusses the variation of effects across countries. This is especially important in the study of these four factors, as a lot of research tries to establish "one result fits all" findings. In this regard, the present study has to disagree with recent aggregate findings such as that democracy causes growth (Acemoglu et al. 2019), democracy does not cause growth (Pozuelo et al. 2016) or trade promotes peace (Hegre et al. 2010). Additionally, this finding touches upon ideological grounds: institutions like the World Bank as well as OECD and other advocates for the liberal cause untiringly and unceasingly insist that trade aids development, reduces conflict and helps to establish democracy (see e.g. Rodriguez and Rodrik (2000)). This paper finds that they are all right, but at the same time, they are all wrong: all depends on the country you are looking at.

This finding of heterogeneous effects is in line with the results of the studies that relax the homogeneous slope parameter assumption (to different degrees),[13] most notably Cervellati et al. (2014). Cervellati et al. (2014) replicate a seminal study, Acemoglu et al. (2008), which finds a robust null-effect of development (income) on democracy once country fixed effects are included. In their replication Cervellati et al. (2014) find that once the effects of income on democracy are allowed to differ across countries there is substantial heterogeneity in slope parameters. They conclude that this heterogeneity (and with it the wide range of point estimates of slope parameters across countries) can provide a plausible reason for the null-result in Acemoglu et al. (2008) or any model assuming homogeneity of slope parameters in this context.

Additionally, hetereogeneity is not limited to the cross-section, but also is important on the within-country level - as discussed in Section 5.5 and shown in Figure 3. This finding is another valid argument against the homogeneous slope parameter assumption on the one hand, but as well for the incorporation of heterogeneity over time.

Certain limitations to the model are worth noting. First, in the interest of parsimony no exogenous control variables are added. The number of possible extensions to this model is limited by page-restrictions only. In future research, an interesting approach would be, for example, to expand on the conflict dimension. The conflict category used in this paper consists of several quite different forms of conflict. This aggregate measure hides the heterogeneous effects of different conflict types (see Kamin (n.d.)). Such conflict types could be added to the model one by one as well as a location variable indicating whether a given country was the location of a conflict in a given year or not. Secondly, for a better understanding of bilateral trade and bilateral conflict, the country-year format employed in this paper could be extended into a country-pair-year format. Bilateral trade and conflict heavily depend on country pair characteristics. Including such information in the model would be an interesting extension.

[13] See for example Rigobon and Rodrik (2005), Narayan et al. (2011).

Appendix

Appendix A

Viva la revolución

A.1 Composition of the Pol Index

This paper analyzes the development of and changes in the governing structures of roughly 140 countries over the past 60 years. Hence, a measure of "system of governance" was needed. Due to its broad chronological and geographical scope the Polity2 Index[1] was the most suitable one for our purpose. It ranges from -10 (strongly autocratic) to 10 (strongly democratic) and is calculated by subtracting a measure of autocracy (AUTOC) from a measure of institutionalized democracy (DEMOC).[2] In this paper democratization thus refers to an "upwards" move on the Polity2 scale.

Examining changes in the system of governance it is essential that the democracy measure is coded on an interval scale. The Polity2 Index, however, is not: an index value of "0" can occur in three different cases. The most intuitive one is if a country's AUTOC score equals its DEMOC score. The second reason why a country might be assigned a Polity2 Index value of "0" in a given year is a transition period. The Polity Project codes years of transition during which "new institutions are planned, legally constituted, and put into effect" resulting in new polities.[3] If a country is undergoing such a transition its Polity2 Index is prorated across the time span of the transition. For example, if it has a transition year in 1990 and index values of "-2" in 1989 and "2" in 1991, then 1990 will be assigned a "0". Both of these codings are suitable for our research since the differences between the Polity2 Index values remain meaningful. However, this is not true for the third case in which a country might be assigned a value of "0". In years in which a total collapse of central political authority occurs the country is assigned a Polity2 Index value of "0". The civil war in Afghanistan, 1992 - 1995, is an example for such a case of "anarchy". The meaning behind this particular "0" is rather different from the others and it renders the differences impossible to interpret. Consequently, the category "complete loss of central political authority" was included for such cases and they were coded separately. The question where to place this category on the scale of measure of "governing structures" is not an easy one to answer since it implies a rating. The Polity2 Index is coded using an "institutional approach" to governing structures. When coding AUTOC and DEMOC the PolityIV Project evaluates a number of factors, namely the openness and competitiveness of executive recruitment, constraints on the chief executive and the competitiveness of political participation. When a country falls into anarchy, its institutions cease to function. Hence, it was positioned at the very bottom of the scale. Another limitation to the Polity2 Index which needed to be changed in our framework is the following: The Polity Project regards years as an interruption during

[1] Retrieved from Marshall et al. (2013).

[2] For a detailed discussion of the components and limitations of the Polity2 index confer to the Polity IV Project Dataset User's Manual, Gurr et al. (2013), pp.13-20.

[3] See Polity IV Project Dataset User's Manual, Gurr et al. (2013), p.19.

which "a country is occupied by foreign powers during war, terminating the old polity, (...) until an independent polity is reestablished" after the end of foreign occupation.[4] Such years are coded as system missing. This is an avoidable loss of information. Since the central political authority of the respective country ceases to function as before, the "interruption" cases are captured in the "complete loss of central political authority" category. In order to differentiate between interruption and anarchy a variable stating whether the central political authority was a) in the hands of the corresponding country, b) collapsed or c) in the hands of foreign powers to our analysis was added. I labelled this redefined Polity2 Index *Pol Index*. In the following $p_{i,t}$ will denote the Pol Index of country i at time t. Then $\Delta p_{i,[t,t+k]} := p_{i,t+k} - p_{i,t}$ is the change in Pol Index for country i between the years t and $t+k$. In the analysis, this paper examines how a number of factors in the base year t are connected to the level of the Pol Index 2, 5 and 10 years later. It focuses on the periods $k = 2, 5, 10$ representing the short, intermediate and long run.

[4]See Polity IV Project Dataset User's Manual, Gurr et al. (2013), p.19.

Appendix B

How (not) to measure democracy

B.1 Factor Analysis of the Polity2 Index components:

Coppedge et al. (2008) carry out a factor analysis for four of the components (parcomp, xropen, xrcomp and xconst) as well as several other measures of democracy. They find evidence for two latent factors - contestation and participation. In their analysis xropen contributes to the inclusiveness/participation factor while the other components are associated with the contestation factor. The results of their analysis might be misleading insofar as Democ, Autoc and Polity Index are not made up of the actual values of the respective components. Due to the intricate weighing and aggregation scheme the indices are sums of the scores on each the components categories. Hence, I created score variables capturing the actual value which is contributed to the Polity Index and carried out an exploratory factor analysis on them. The results for the factor analysis using the score variables are displayed in Table B.1. One latent factor explains 95.86% of the variation the Polity Index's components - contestation (or, as Munck and Verkuilen put it: the procedure by which the executive office is filled). These findings in combination with the absence of any suffrage/inclusiveness requirement suggest that the Polity Index in fact is a measure of political contestation rather than democracy - even if one embraces the minimalist democracy definition with contestation and participation.

Table B.1: Factor analysis results for the score variables of Polity's five components

Factor	Eigenvalue	Proportion
1	3.56	0.9586
2	0.36	0.0981
3	-0.01	-0.0038
N=	17,228	

Variable: scores of...	Factor Loadings (Factor 1)	Uniqueness	KMO
...xrcomp	0.8969	0.1955	0.7771
...xropen	0.8337	0.3050	0.7778
...xconst	0.8039	0.3537	0.9209
...parcomp	0.8504	0.2768	0.7833
...parreg	0.8314	0.3087	0.7925
		Overall:	0.8055

B.2 Checklist of questions and respectives scores for the components of the Freedom House Index

Table B.2: Checklist for the Freedom House Index. Source: Freedom House Methodology Website, Freedom House (2017b)

	Score	Political Rights	Score	Civil Liberties
	0–12 points	**A. Electoral Process** 1. Is the head of government or other chief national authority elected through free and fair elections? 2. Are the national legislative representatives elected through free and fair elections? 3. Are the electoral laws and framework fair?	0–16 points	**D. Freedom of Expression and Belief** **1.** Are there free and independent media and other forms of cultural expression? (*Note*: In cases where the media are state controlled but offer pluralistic points of view, the survey gives the system credit.) **2.** Are religious institutions and communities free to practice their faith and express themselves in public and private? **3.** Is there academic freedom, and is the educational system free of extensive political indoctrination? **4.** Is there open and free private discussion?
	0–16 points	**B. Political Pluralism and Participation** 1. Do the people have the right to organize in different political parties or other competitive political groupings of their choice, and is the system open to the rise and fall of these competing parties or groupings? 2. Is there a significant opposition vote and a realistic opportunity for the opposition to increase its support or gain power through elections? 3. Are the people's political choices free from domination by the military, foreign powers, totalitarian parties, religious hierarchies, economic oligarchies, or any other powerful group? 4. Do cultural, ethnic, religious, or other minority groups have full political rights and electoral opportunities?	0–12 points	**E. Associational and Organizational Rights** **1.** Is there freedom of assembly, demonstration, and open public discussion? **2.** Is there freedom for nongovernmental organizations? (*Note*: This includes civic organizations, interest groups, foundations, etc., with an emphasis on those engaged in human rights– and governance-related work.) **3.** Are there free trade unions and peasant organizations or equivalents, and is there effective collective bargaining? Are there free professional and other private organizations?
	0–12 points	**C. Functioning of Government** **1.** Do the freely elected head of government and national legislative representatives determine the policies of the government? **2.** Is the government free from pervasive corruption? **3.** Is the government accountable to the electorate between elections, and does it operate with openness and transparency?	0–16 points	**F. Rule of Law** **1.** Is there an independent judiciary? **2.** Does the rule of law prevail in civil and criminal matters? Are police under direct civilian control? **3.** Is there protection from political terror, unjustified imprisonment, exile, or torture, whether by groups that support or oppose the system? Is there freedom from war and insurgencies? **4.** Do laws, policies, and practices guarantee equal treatment of various segments of the population?
	(0-4 points) (–4 to 0 points)	**Additional Discretionary Political Rights Questions** A. For traditional monarchies that have no parties or electoral process, does the system provide for genuine, meaningful consultation with the people, encourage public discussion of policy choices, and allow the right to petition the ruler? B. Is the government or occupying power deliberately changing the ethnic composition of a country or territory so as to destroy a culture or tip the political balance in favor of another group? **Note:** For additional discretionary question A, a score of 1 to 4 may be added, as applicable, while for discretionary question B, a score of 1 to 4 may be subtracted, as applicable (the worse the situation, the more points may be subtracted).	0–16 points	**G. Personal Autonomy and Individual Rights** 1. Do individuals enjoy freedom of travel or choice of residence, employment, or institution of higher education? 2. Do individuals have the right to own property and establish private businesses? Is private business activity unduly influenced by government officials, the security forces, political parties/organizations, or organized crime? 3. Are there personal social freedoms, including gender equality, choice of marriage partners, and size of family? 4. Is there equality of opportunity and the absence of economic exploitation?
Total	0 - 40 points	With the two discretionary questions the highest possible score remains 40, but the lowest possible score is -4.	0 - 60 points	

B.3 Setup of V-Dem's Electoral Democracy Index (Polyarchy)

Table B.3: Setup of the Electoral Democracy (Polyarchy) Index. The information displayed here is gathered from the V-Dem Codebook, Coppedge et al. (2017a), p.49 - 59 and p. 435-436.

Electoral Democracy Index, **v2x_polyarchy,**

Question: To what extent are rulers responsive to citizens?(...) [It is] achieved through electoral competition for the electorate's approval under circumstances when suffrage is extensive; political and civil society organizations can operate freely; elections are clean and not marred by fraud or systematic irregularities; and elections affect the composition of the chief executive of the country. In between elections, there is freedom of expression and an independent media capable of presenting alternative views on matters of political relevance.

The aggregation is done at the level of Dahl's sub-components (with the one exception of the non-electoral component). The index is aggregated using this formula:

$$v2x\_polyarchy = \frac{1}{2}v2x\_api + \frac{1}{2}v2x\_mpi$$

Mid-Level Indices:

Additive Polyarchy Index, **v2x_api**

Question: To what extent is the electoral principle of democracy achieved?

Clarification: The electoral principle of democracy seeks to achieve responsiveness and accountability between leaders and citizens through the mechanism of competitive elections. This is presumed to be achieved when suffrage is extensive; political and civil society organizations can operate freely; elections are clean and not marred by fraud or systematic irregularities; and the chief executive of a country is selected (directly or indirectly) through elections.

$$v2x\_api = \frac{1}{4}v2x\_frassoc\_thick + \frac{1}{4}v2xel\_frefair + \frac{1}{4}v2x\_freexp\_thick + \frac{1}{4}\left(\frac{1}{2}v2x\_elecoff + \frac{1}{2}v2x\_suffr\right)$$

Multiplicative polyarchy index, **v2x_mpi**

$$v2x_{mpi} = v2x_{frassoc_{thick}} \cdot v2xel_{frefair} \cdot v2x_{freexp_{thick}} \cdot \ldots \ldots \cdot v2x\_elecoff \cdot v2x\_suffr$$

Low-Level Indices:

Freedom of expression index , **v2x_freexp_thick**	Freedom of association index, **v2x_frassoc_thick**	Clean elections index, **v2xel_frefair**	Elected official index, **v2x_elecoff**	Share of population with suffrage, **v2x_suffr**
Type A (factual), B and C (expert coded) variables				
Government censorship effort of media, Harassment of journalists, Media self-censorship, Media bias, Print/broadcast media critical, Print/broadcast media perspectives; Freedom of discussion for men; Freedom of discussion for women; Freedom of academic and cultural expression)	Party ban, Barriers to parties, Opposition parties autonomy, Elections multiparty, civil society entry and exit, civil society repression	EMB autonomy, EMB capacity, Election voter registry, Election vote buying, Election other voting irregularities, Election government intimidation, Election other electoral violence, Election free and fair	Legislature bicameral, Lower chamber elected, Upper chamber elected, Legislature dominant chamber, head of state(HOS) selection by legislature in practice, HOS appointment in practice, HOG selection by legislature in practice, HOG appointment in practice, HOS appoints cabinet in practice, HOG appoints cabinet in practice, HOS dismisses ministers in practice, HOG dismisses ministers in practice, HOS = HOG? Chief executive appointment by upper chamber, Chief executive appointment by upper chamber explicit approval	Percent of population with suffrage

B.4 Derivation of the Polyarchy Index' rate of influence

$$\text{democracy index} = \frac{1}{4} \cdot (\text{polyarchy}^{1.6} + \text{component index}) + \frac{1}{2} \cdot \text{polyarchy}^{1.6} \cdot \text{component index} \tag{B.1}$$

The polyarchy index influences each high level democracy index to the power of 1.6. The intuition behind this rate is explained in the V-Dem Methodology V7 paper, Coppedge et al. (2017e), p.10: "when a country has a polyarchy score of .5 (in practice, this is a threshold on the Electoral Democracy Index beyond which countries tend to be considered electoral democracies in a minimal sense) and its HPC[1] is at its maximum (1), the high level index score should be .5".

$$0.5 = \frac{1}{4} \cdot (0.5^x + 1) + \frac{1}{2} \cdot 0.5^x \cdot 1 \rightarrow x \approx 1.6 \tag{B.2}$$

This benchmark case is shown in equation 1. Solving for x yields a rate of close to 1.6. Intuitively, setting a higher rate of influence for polyarchy than any of the other component indices reflects a notion of democracy being a question of kind before one of degree. The principles of contestation and participation should be satisfied to a certain degree before further aspects of democracy can be employed to distinguish between regime types.

B.5 Summary statistics for the V-Dem democracy measures

	Summary statistics for...					
Democracy	...all observations available per index					
Index:	Obs.	Mean	Median	St. D.	Min	Max
Polyarchy (Electoral)	17,036	0.318	0.206	0.279	0.009	0.947
Liberal	17,035	0.260	0.151	0.246	0.009	0.916
Participatory	17,035	0.192	0.105	0.193	0.005	0.814
Deliberative	17,035	0.209	0.068	0.262	0.000	0.913
Egalitarian	17,036	0.242	0.148	0.232	0.011	0.890

	...the trunk dataset					
Index:	Obs.	Mean	Median	St. D.	Min	Max
Polyarchy (Electoral)	6,546	0.455	0.406	0.287	0.014	0.947
Liberal	6,546	0.354	0.260	0.279	0.010	0.916
Participatory	6,546	0.279	0.212	0.210	0.007	0.814
Deliberative	6,546	0.330	0.236	0.294	0.001	0.913
Egalitarian	6,546	0.356	0.260	0.249	0.032	0.890

Table B.4: Summary Statistics for V-Dem democracy measures for all observations available (left) and observations in the trunk dataset (right).

[1] HPC refers to High Principle Component (here: component index).

B.6 Regression of difference between democracy measures on "hypothetical democracy"

Dep. Var.:	Polity2 - Polyarchy Coeff.	Polity2 - FHI Coeff	FHI-Polyarchy Coeff
hypothetical democracy	0.2326***	0.0857***	0.1469***
	(0.0055)	(0.0064)	(.0049)
constant	-0.0040	0.0120***	-0.0241***
	(0.0033)	(0.0038)	(.0029)
R^2	0.2175	0.0266	0.1252
N	6,546	6.546	6,546

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.10$

Table B.5: Regression results for regression of difference between democracy measures on "hypothetical democracy". Pooled OLS, standard deviation in parenthesis below.
Independent variable: "hypothetical democracy":= $\frac{(Polity2+FHI+Polyarchy)}{3}$

B.7 Description of the democracy measures' distribution

Figure B.1: Histogram of the normalized democracy measures in trunk dataset

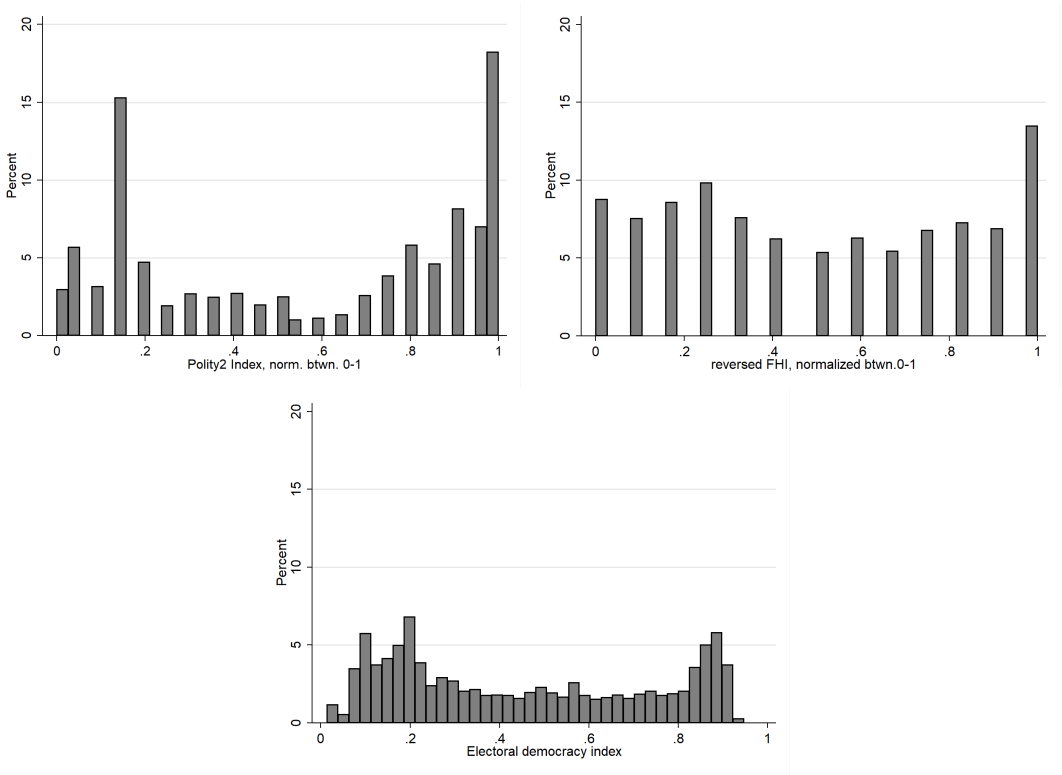
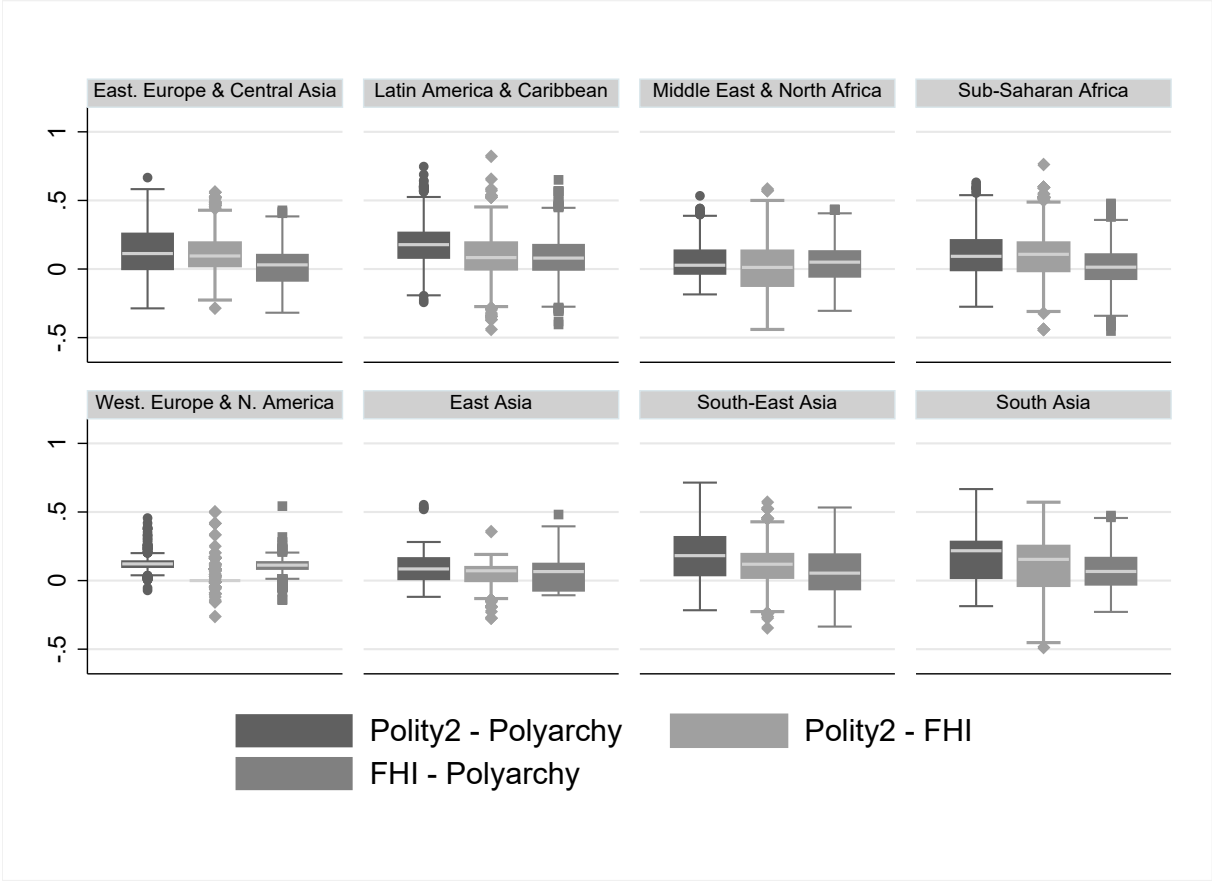


Table B.6: Percentiles for Polity2 (normalized between 0-1), Polyarchy and FHI (reversed and normalized between 0-1) in trunk dataset

	Polity2		Polyarchy		FHI	
	Percentiles	Smallest	Percentiles	Smallest	Percentiles	Smallest
1%	0	0	0.0257	0.0140	0	0
5%	0.05	0	0.0858	0.0140	0	0
10%	0.1	0	0.1069	0.0140	0.0833	0
25%	0.15	0	0.1911	0.0140	0.2500	0
50%	0.7		0.4061		0.5	
		Largest		Largest		Largest
75%	0.95	1	0.7380	0.9335	0.8333	1
90%	1	1	0.8725	0.9357	1	1
95%	1	1	0.8912	0.9448	1	1
99%	1	1	0.9103	0.9471	1	1

Figure B.2: Boxplot of differences in democracy measures by geopolitical region



B.8 Examining differences in regional codings and rankings

To examine the distribution of the differences between the index values a boxplot by region is provided in Figure B.2. Polity2 mostly assigns higher values than the other indices, followed by the FHI. The inter quartile range of the differences between Polity2 and the other two indices is non-negative for all regions exept MENA and South Asia (for Polity2 - FHI). There is a very high level of agreement between all indices for Western Europe and North America (the differences are mostly positive but close to 0). All indices also display a high level of agreement for East Asia. The picture for differences between FHI and Polyarchy is mixed, although the FHI mostly assigns larger values than the Polyarchy Index (median differences are positive for all regions, interquartile range non-negative for Latin America & the Caribbean, Western Europe & North America).

B.8.1 Ranking

While the differences discussed above provide some information regarding the general "austereness" of each index they do not give any information on how the countries are coded in each year with respect to each other. When ranking all countries according to their democracy index value in a given year is each country ranked consistently accross the three measures? In the following the country rankings will be compared across the three democracy measures. Note, that this exercise has its limitations due to the uncertainty embedded in each measure. For an excellent discussion of this see Høyland et al. (2012). For each year, all countries were ranked according to their democracy index values obtaining a rank in Polity2 values, a rank in FHI values and a rank in Polyarchy values. The regional average democracy values from Figure 3.5

Figure B.3: Regional average ranks based on the yearly values coded by each of the democracy indices

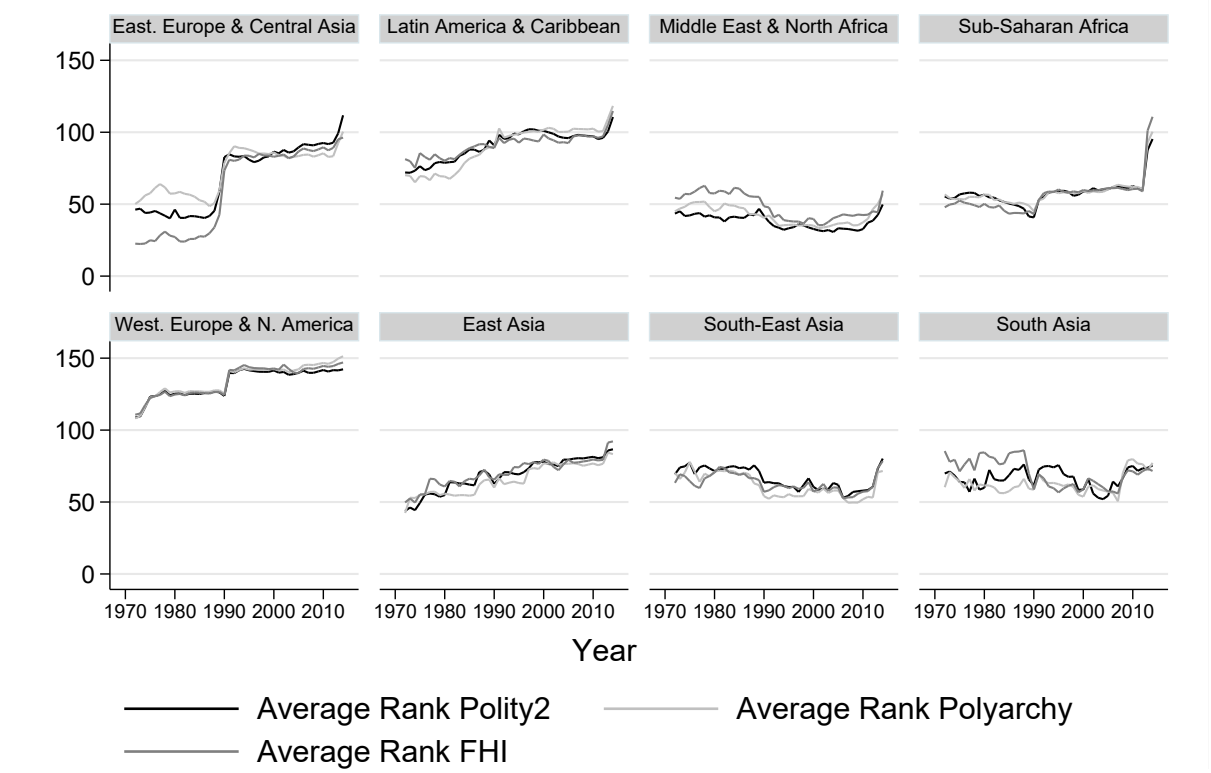


Table B.7: Kendall's τ_b (upper right side) and Spearman's ρ (bottom left side)

	Rank Polity2	Rank Polyarchy	Rank FHI
Rank Polity2	-	0.6903***	0.6909***
Rank Polyarchy	0.8735***	-	0.7303***
Rank FHI	0.8729***	0.8993***	-

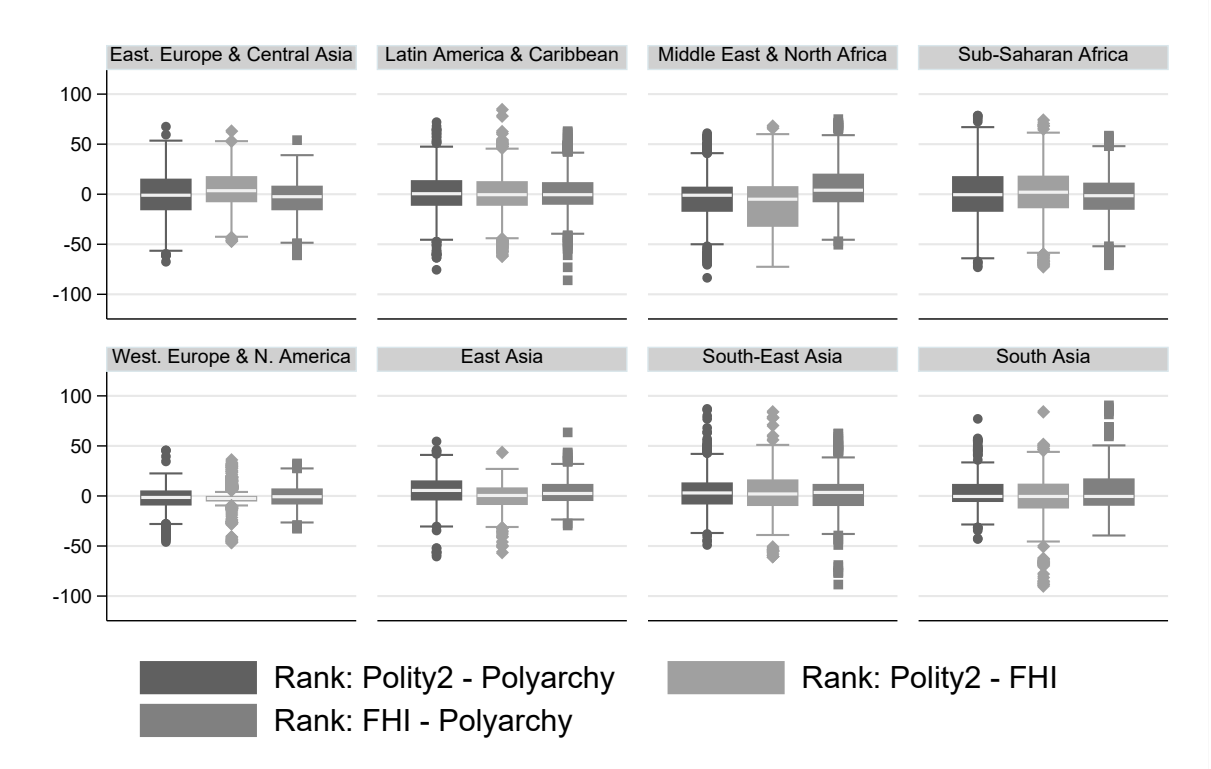
$^{***}p < 0.01$,$^{**}p < 0.05$, $^{*}p < 0.10$

are thus reproduced as regional average ranks in Figure B.3. A rank of 1 corresponds to the lowest possible democracy index value. The ranks are coded keeping the overall sum constant. Note that the ranking should only be compared within single years but not over time as one country's ranking can change when other countries are coded as democratic.

Western European & North American, Sub-Saharan African and East Asian countries are ranked very similarly by the indices. The difference in values for Eastern Europe & Central Asia translates into the most pronounced difference in rankings. Hence, in research frameworks in which countries of that time period and region are considered it is very likely that the results will vary with the index used. Therefore, using the FHI in such cases is inadvisable. Further notable differences in rankings occur before 1990 in the MENA region and South Asia. In these cases, FHI assigns higher average democracy values than the other two indices.

Figure B.4 displays a boxplot of the differences between index rankings by region. The difference in ranks has the highest variability for Polity2-Polyarchy in Sub-Saharan Africa. Figure B.4 shows that the medians for the difference in ranges are in most cases close to and the interquartile ranges centered around 0. This suggests that the differences in index values do not translate into difference in ranks.

Figure B.4: Boxplot of differences in yearly country rankings between indices by region



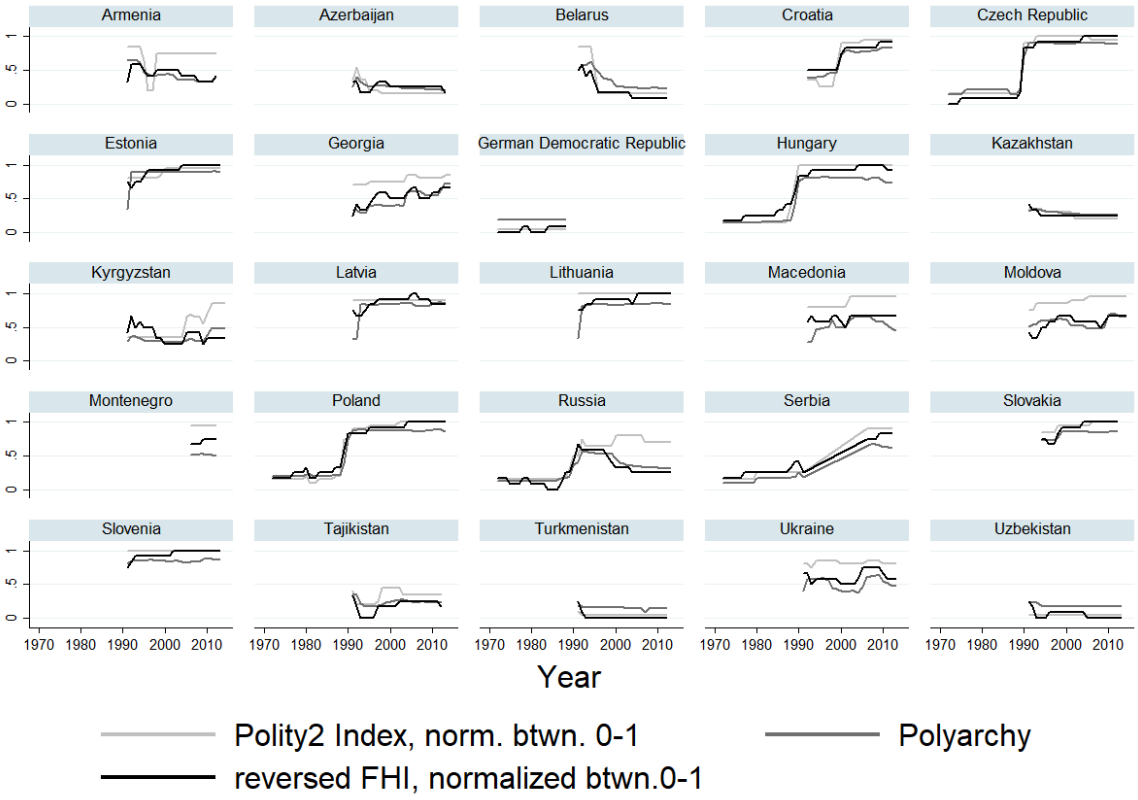
Lastly, the similarities of the rankings are compared using Kendall's τ[2] and Spearman's ρ. A τ or a ρ close to 1 implies a strong association between the rankings. Spearman's ρ "accepts" small differences in ranking from time to time and is very sensitive to large errors (even if they occur only once). Kendall's τ on the other hand is relatively insensitive to large errors occurring just once. The rank correlation coefficients are displayed in Table B.7. Both, Kendall's τ and Spearman's ρ, point to a very strong agreement with respect to the ranking for all variables. The nullhypothesis of independent rankings is rejected in all pairwise comparisons. In sum, the vast majority of country years is ranked consistently across all indices.

[2]To include ties we used τ_b .

B.9 Eastern Europe and Central Asia (EECA)

Figure B.5 displays the country codings by Polity2, FHI and Polyarchy for each country in the EECA region with more than 7 years of data available in the trunk dataset (Kosovo and Bosnia and Herzegovina did not meet this criterion and were thus left out, see Table B.10).

Figure B.5: Democracy coding by Polity2, FHI and Polyarchy for selected countries in Eastern Europe & Central Asia



B.10 Countries and their respective geopolitical regions

Table B.8: Countries and their respective geopolitical regions as coded in section *Comparison*.

Region		Countries
1	East. Europe & Central Asia (31 countries)	Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegowina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, German Democratic Republic, Hungary, Kazakhstan, Kosovo, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Mongolia, Montenegro, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan
2	Latin America & Caribbean (24 countries)	Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Trinidad and Tobago, Uruguay, Venezuela
3	Middle East & North Africa (18 countries)	Algeria, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, South Yemen, Syria, Tunisia, Turkey, Yemen
4	Sub-Saharan Africa (47 countries)	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Republic of the Congo, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
5	West. Europe & North America (21 countries)	Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States
6	East Asia (5 countries)	China, Japan, North Korea, South Korea, Taiwan
7	South-East Asia (14 countries)	Burma/Myanmar, Cambodia, Democratic Republic of Vietnam, Fiji, Indonesia, Laos, Malaysia, Papua New Guinea, Philippines, Republic of Vietnam, Singapore, Solomon Islands, Thailand, Timor-Leste
8	South Asia (7 countries)	Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, Sri Lanka

B.11 Slow changing level of democracy in Polity2 Index

When using the Polity2 Index in time series and/or cross section models one factor to keep in mind is the slow changing nature of the level of democracy. In most years the level of a country's previous democracy index is the best predictor for its' current value. Gleditsch and Ward (1997) examined these changes in the Polity II Data (one of Polity IV's predecessors) with the help of Markov transition matrices. They show that much of the variation in the polity index is cross sectional rather than temporal. Table B.9 examines the variation in the Polity2 Index more closely.

Table B.9: Examining variation in the Polity2 Index

	Overall		Between		Within
polity2	Freq.	Percent	Freq.	Percent	Percent
-10	1354	7.97	33	17.01	42.54
-9	1144	6.73	62	31.96	23.75
-8	512	3.01	50	25.77	13.32
-7	1,884	11.09	107	55.15	25.44
-6	1297	7.63	82	42.27	13.80
-5	586	3.45	64	32.99	8.92
-4	715	4.21	65	33.51	12.41
-3	1,166	6.86	67	34.54	15.27
-2	324	1.91	48	24.74	9.03
-1	540	3.18	48	24.74	10.93
0	415	2.44	69	35.57	9.66
1	379	2.23	37	19.07	10.77
2	464	2.73	52	26.80	9.72
3	282	1.66	37	19.07	6.84
4	535	3.15	49	25.26	13.73
5	441	2.60	58	29.90	11.08
6	547	3.22	58	29.90	14.62
7	549	3.23	61	31.44	12.59
8	773	4.55	65	33.51	15.25
9	645	3.80	48	24.74	18.53
10	2,440	14.36	43	22.16	46.71
Total	16,992	100.00	1,203 (n=194)	620.10	16.13

For the overall part the unit of observation is a country-year. There are 1,354 observations in which a country in a given year obtained a polity index value of -10. In the between part the unit of observation is a country instead of a country-year; 33 of the countries ever had a Polity2 value of -10 and a total of 1203 countries was categorized. Due to the fact that the data only includes 194 countries, it follows that some countries switched between the categories. The within percent indicates the percentage of the time a country has the specified Polity2 value. Conditional on a country ever having a Polity2 value of -10, 42.54% of that same country's observations have the same index value. Interestingly, this percentage increases for both "high" democracies and autocracies. Conditional on a country ever having obtained a Polity2 value of 10, 46.71% of that country's observations have the same index value. This fraction is much smaller (around 10%) for Polity2 values between -2 and 2.

B.12 Countries and years included in the trunk dataset

Table B.10: Countries (A-F) and years included in the trunk dataset

country	first year	last year	# years in sample	# of missing years
Afghanistan	1972	2015	21	23
Albania	1972	2015	44	0
Algeria	1972	2015	44	0
Angola	1975	2015	41	0
Argentina	1972	2015	44	0
Armenia	1991	2015	25	0
Australia	1972	2015	44	0
Austria	1972	2015	44	0
Azerbaijan	1991	2015	25	0
Bangladesh	1972	2015	44	0
Belarus	1991	2015	25	0
Belgium	1972	2015	44	0
Benin	1972	2015	44	0
Bhutan	1972	2015	44	0
Bolivia	1972	2015	44	0
Bosnia and Herzegovina	1993	1994	2	0
Botswana	1972	2015	44	0
Brazil	1972	2015	44	0
Bulgaria	1972	2015	44	0
Burkina Faso	1972	2015	44	0
Burma/Myanmar	1972	2014	43	0
Burundi	1972	2015	44	0
Cambodia	1972	2015	35	9
Cameroon	1972	2015	44	0
Canada	1972	2015	44	0
Cape Verde	1975	2015	41	0
Central African Republic	1972	2015	44	0
Chad	1972	2015	44	0
Chile	1972	2015	44	0
China	1972	2015	44	0
Colombia	1972	2015	44	0
Comoros	1975	2015	41	0
Costa Rica	1972	2015	44	0
Croatia	1991	2015	25	0
Cuba	1972	2015	44	0
Cyprus	1972	2015	44	0
Czech Republic	1972	2015	44	0
Democratic Republic of Congo	1972	2015	44	0
Democratic Republic of Vietnam	1972	2015	44	0
Denmark	1972	2015	44	0
Djibouti	1977	2015	39	0
Dominican Republic	1972	2015	44	0
Ecuador	1972	2015	44	0
Egypt	1972	2015	44	0
El Salvador	1972	2015	44	0
Equatorial Guinea	1972	2015	44	0
Eritrea	1993	2015	23	0
Estonia	1991	2015	25	0
Ethiopia	1972	2015	44	0
Fiji	1972	2015	44	0
Finland	1972	2015	44	0
France	1972	2015	44	0

Table B.11: Countries (G-N) and years in trunk dataset continued

country	first year	last year	# years in sample	# of missing years
Gabon	1972	2015	44	0
Gambia	1972	2015	44	0
Georgia	1991	2015	25	0
German Democratic Republic	1972	1988	17	0
Germany	1972	2015	44	0
Ghana	1972	2015	44	0
Greece	1972	2015	44	0
Guatemala	1972	2015	44	0
Guinea	1972	2015	44	0
Guinea-Bissau	1974	2015	42	0
Guyana	1972	2015	44	0
Haiti	1972	2015	44	0
Honduras	1972	2015	44	0
Hungary	1972	2015	44	0
India	1972	2015	44	0
Indonesia	1972	2015	44	0
Iran	1972	2015	44	0
Iraq	1972	2015	37	7
Ireland	1972	2015	44	0
Israel	1972	2015	44	0
Italy	1972	2015	44	0
Ivory Coast	1972	2015	44	0
Jamaica	1972	2015	44	0
Japan	1972	2015	44	0
Jordan	1972	2015	44	0
Kazakhstan	1991	2015	25	0
Kenya	1972	2015	44	0
Kosovo	2009	2015	7	0
Kuwait	1972	2015	43	1
Kyrgyzstan	1991	2015	25	0
Laos	1972	2015	44	0
Latvia	1991	2015	25	0
Lebanon	1972	2015	29	15
Lesotho	1972	2015	44	0
Liberia	1972	2015	44	0
Libya	1972	2015	44	0
Lithuania	1991	2015	25	0
Macedonia	1992	2015	24	0
Madagascar	1972	2015	44	0
Malawi	1972	2015	44	0
Malaysia	1972	2015	44	0
Mali	1972	2015	44	0
Mauritania	1972	2015	44	0
Mauritius	1972	2015	44	0
Mexico	1972	2015	44	0
Moldova	1991	2015	25	0
Mongolia	1972	2015	44	0
Montenegro	2006	2015	10	0
Morocco	1972	2015	44	0
Mozambique	1978	2015	38	0
Namibia	1990	2015	26	0
Nepal	1972	2015	44	0
Netherlands	1972	2015	44	0
New Zealand	1972	2015	44	0
Nicaragua	1972	2015	44	0
Niger	1972	2015	44	0
Nigeria	1972	2015	44	0
North Korea	1972	2015	44	0
Norway	1972	2015	44	0

Table B.12: Countries (O-Z) and years in trunk dataset continued

country	first year	last year	# years in sample	# of missing years
Oman	2000	2015	16	0
Pakistan	1972	2015	44	0
Panama	1972	2015	44	0
Papua New Guinea	1975	2015	41	0
Paraguay	1972	2015	44	0
Peru	1972	2015	44	0
Philippines	1972	2015	44	0
Poland	1972	2015	44	0
Portugal	1972	2015	44	0
Qatar	1972	2015	44	0
Republic of Vietnam	1973	1975	3	0
Republic of the Congo	1972	2015	44	0
Romania	1972	2015	44	0
Russia	1972	2015	44	0
Rwanda	1972	2015	44	0
Saudi Arabia	1972	2015	44	0
Senegal	1972	2015	44	0
Serbia	1972	2015	30	14
Sierra Leone	1972	2015	44	0
Singapore	1972	2015	44	0
Slovakia	1994	2015	22	0
Slovenia	1991	2015	25	0
Solomon Islands	1978	2015	37	1
Somalia	1972	2015	43	1
South Africa	1972	2015	44	0
South Korea	1972	2015	44	0
South Sudan	2012	2015	4	0
South Yemen	1972	1989	18	0
Spain	1972	2015	44	0
Sri Lanka	1972	2015	44	0
Sudan	1972	2011	40	0
Suriname	1975	2015	41	0
Swaziland	1972	2015	44	0
Sweden	1972	2015	44	0
Switzerland	1972	2015	44	0
Syria	1972	2015	44	0
Taiwan	1972	2015	44	0
Tajikistan	1991	2015	25	0
Tanzania	1972	2015	44	0
Thailand	1972	2015	44	0
Timor-Leste	2002	2015	14	0
Togo	1972	2015	44	0
Trinidad and Tobago	1972	2015	44	0
Tunisia	1972	2015	44	0
Turkey	1972	2015	44	0
Turkmenistan	1991	2015	25	0
Uganda	1972	2015	43	1
Ukraine	1991	2015	25	0
United Kingdom	1972	2015	44	0
United States	1972	2015	44	0
Uruguay	1972	2015	44	0
Uzbekistan	1991	2015	25	0
Venezuela	1972	2015	44	0
Yemen	1972	2015	43	1
Zambia	1972	2015	44	0
Zimbabwe	1972	2015	44	0

Appendix C

'Tis but thy name that is my enemy

The tables in this Appendix are also online (see Boese and Kamin, 2018a, 2018b, which provide a much more detailed overview of which countries and years actually contain nonmissing values in each of the datasets).

C.1 Democracy Datasets Comparison

See Boese and Kamin (2018a) for a very detailed listing of all countries and their respective time series covered. Countries for which only the names/labels differ are listed in Table C.1 (that is, countries of inconsistency type 3, reason 1.) In the worksheet "Overview" (Boese and Kamin, 2018a), these countries are highlighted in grey.

V-Dem Version 8	Polity IV, Version 2016
Bosnia and Herzegovina	Bosnia
Burma/Myanmar	Myanmar (Burma)
Democratic Republic of Congo	Congo Kinshasa
German Democratic Republic	Germany East
North Korea	Korea North
Piedmont-Sardinia	Sardinia
Republic of Vietnam	Vietnam South
Republic of the Congo	Congo Brazzaville
Slovakia	Slovak Republic
South Korea	Korea South
South Yemen	Yemen South
United Arab Emirates	UAE
United States of America	United States
Würtemberg	Wuerttemburg

Table C.1: Countries for which only the names/labels differ (democracy datasets)

Countries for which the underlying entity has no perfect match in the other dataset are listed in Table C.2. A "perfect match" refers to a counterpart in terms of names and years (and potentially borders). This includes countries of inconsistency types 1 and 3. Countries representing the same or similar historical units are grouped.

Countries unmergable due to name and time inconsistencies are listed in Table C.3. This includes countries of inconsistency type 3. Note: # obs=number of observations; N=total number of available observations in data; missing=number of missing years/observations for given country between its first and last year.

V-Dem Version 8	Polity IV, Version 2016
Barbados	
Brunswick	
Colombia	Colombia
	Gran Colombia
Czech Republic	Czech Republic
	Czechoslovakia
Democratic Republic of Vietnam	Vietnam North
	Vietnam
German Democratic Republic	Germany East
Germany	Germany
	Prussia
	Germany West
Guatemala	United Province of CA (Central America)
Hamburg	
Hanover	
Hesse-Darmstadt	
Hesse-Kassel	
Hong Kong	
Iceland	
Ivory Coast	Ivory Coast
	Cote D'Ivoire
Maldives	
Mecklenburg Schwerin	
Nassau	
Oldenburg	
	Orange Free State
Palestine/British Mandate	
Palestine/Gaza	
Palestine/West Bank	
Russia	USSR
Saxe-Weimar-Eisenach	
Serbia	Serbia
	Serbia and Montenegro
	Yugoslavia
Seychelles	
Somaliland	
South Korea	Korea South
	Korea
South Sudan	South Sudan
Sudan	Sudan
	Sudan-North
São Tomé and Príncipe	
Timor-Leste	Timor Leste
	East Timor
Vanuatu	
Yemen	Yemen
	Yemen North
Zanzibar	

Table C.2: Countries for which the underlying entity has no perfect match in the other dataset (Democracy Datasets)

V-Dem Version 8, 201 countries					Polity IV, Version 2016, 195 countries				
	year		# obs. in data			year		# obs. in data	
country	first	last	N	missing	country	first	last	N	missing
Bosnia and Herzegovina	1992	2017	26	0	Bosnia	1992	2016	25	0
					Yugoslavia	1921	2002	83	-1
Colombia	1789	2017	229	0	Colombia	1832	2016	185	0
					Gran Colombia	1821	1832	12	0
Czech Republic	1918	2017	100	0	Czech Republic	1993	2016	24	0
					Czechoslovakia	1918	1992	75	0
Democratic Republic of Vietnam	1945	2017	73	0	Vietnam North	1954	1976	23	0
					Vietnam	1976	2016	41	0
Germany	1789	2017	225	4	Germany	1868	2016	105	44
					Prussia	1800	1867	68	0
					Germany West	1945	1990	46	0
Ivory Coast	1900	2017	118	0	Ivory Coast	1960	2015	56	0
					Cote D'Ivoire	2016	2016	1	0
Russia	1789	2017	229	0	Russia	1800	2016	148	69
					USSR	1922	1991	70	0
Serbia	1804	2017	213	1	Serbia	1830	2016	102	85
					Serbia and Montenegro	2003	2006	4	0
South Korea	1789	2017	229	0	Korea South	1948	2016	69	0
					Korea	1800	1910	111	0
Sudan	1900	2017	118	0	Sudan	1956	2011	56	0
					Sudan-North	2011	2016	6	0
South Yemen	1900	1990	91	0	Yemen South	1967	1990	24	0
Yemen	1789	2017	162	67	Yemen	1990	2016	27	0
					Yemen North	1918	1990	73	0
Timor-Leste	1900	2017	118	0	Timor Leste	2016	2016	1	0
					East Timor	2002	2015	14	0

Table C.3: Countries unmergeable due to name and time inconsistencies (democracy datasets)

C.2 Economic Datasets Comparison

Table C.4 is a listing of unmergeable names/labels in the UN Comtrade and WDI datasets, due to inconsistency type 3, and shows a large share of countries with high export levels (Boese and Kamin, 2018b, contains the list sorted by total exports; worksheet "Unmergable Outliers Comtrade". The spreadsheet also provides a list of country groups/regions which were not included in the comparison; worksheet "Disregarded Country Groups").

Tables C.5 and C.6 show countries for which the underlying entity has no perfect match in the other dataset. A "perfect match" refers to a counterpart in terms of names and years (and potentially borders). This includes countries of inconsistency types 1 and 3. Countries representing the same or similar historical units are grouped.

Table C.7 shows countries unmergable due to name and time inconsistencies. This includes countries of inconsistency type 3 (N=total number of available observations in data).

UN Comtrade exports	WDI trade openness
Bolivia (Plurinational State of)	Bolivia
Bosnia Herzegovina	Bosnia and Herzegovina
Cabo Verde	Cape Verde
Cayman Isds	Cayman Islands
Central African Rep.	Central African Republic
China, Hong Kong SAR	Hong Kong
China, Macao SAR	Macao SAR, China
Congo	Republic of the Congo
Czechia	Czech Republic
Côte d'Ivoire	Ivory Coast
Dem. Rep. of the Congo	Democratic Republic of Congo
Dominican Rep.	Dominican Republic
FS Micronesia	Micronesia, Fed. Sts.
Faeroe Isds	Faroe Islands
Gambia	The Gambia
Lao People's Dem. Rep.	Laos
Myanmar	Burma/Myanmar
Rep. of Korea	South Korea
Rep. of Moldova	Moldova
Russian Federation	Russia
Saint Kitts and Nevis	St. Kitts and Nevis
Saint Lucia	St. Lucia
Saint Vincent and the Grenadines	St. Vincent and the Grenadines
Sao Tome and Principe	São Tomé and Príncipe
Solomon Isds	Solomon Islands
TFYR of Macedonia	Macedonia
Turks and Caicos Isds	Turks and Caicos Islands
US Virgin Isds	Virgin Islands (U.S.)
USA	United States of America
United Rep. of Tanzania	Tanzania
Viet Nam	Vietnam
Yemen	Yemen, Rep.

Table C.4: Countries for which the names/labels differ (economic datasets)

C.3 Conflict Dataset

Tables C.8 to C.12 provide a comparison of country coding units in the UCDP/PRIO Armed Conflict dataset 18.1 to the coding units supplied in the relevant code book. Countries with inconsistent labels are written in blue; countries which only exist in the dataset but not in code book are written in red.

UN Comtrade exports	WDI trade openness
	American Samoa
Belgium	Belgium
Belgium-Luxembourg	
	British Virgin Islands
	Channel Islands
Cook Isds	
	Curacao
Czechia	Czech Republic
Czechoslovakia	
East and West Pakistan	
	Equatorial Guinea
Ethiopia	Ethiopia
Fmr Ethiopia	
Fmr Tanganyika	
Fmr Yugoslavia	
French Guiana	
Germany	Germany
Fmr Dem. Rep. of Germany	
Fmr Fed. Rep. of Germany	
	Gibraltar
Guadeloupe	
	Guam
India	India
India, excl. Sikkim	
	Isle of Man
	Kosovo
	Liechtenstein
	Marshall Islands
Mayotte	
	Monaco
Montserrat	
	Nauru
Neth. Antilles	
Neth. Antilles and Aruba	
Niue	
	North Korea
	Northern Mariana Islands
Panama	Panama
Fmr Panama, excl.Canal Zone	
Peninsula Malaysia	
	Puerto Rico
Réunion	

Table C.5: Countries for which the underlying entity has no perfect match in the other dataset, A-R (economic datasets)

UN Comtrade exports	WDI trade openness
Sabah	
Saint Kitts, Nevis and Anguilla	
Saint Pierre and Miquelon	
	San Marino
Serbia and Montenegro	
	Sint Maarten (Dutch part)
State of Palestine	
	St. Martin (French part)
Sudan	Sudan
Fmr Sudan	
	South Sudan
USA	United States of America
USA (before 1981)	
	Uzbekistan
Viet Nam	Vietnam
Fmr Rep. of Vietnam	
Yemen	Yemen, Rep.
Fmr Arab Rep. of Yemen	
	West Bank and Gaza

Table C.6: Countries for which the underlying entity has no perfect match in the other dataset, S-Z (economic datasets)

UN Comtrade exports years available (coded and non-missing)				WDI tradeopenness years available (coded and non-missing)			
country	year		N	country	year		N
	first	last			first	last	
Belgium	1999	2017	19	Belgium	1960	2016	57
Belgium-Luxembourg	1962	1998	30				
Bosnia Herzegovina	2003	2017	15	Bosnia and Herzegovina	1994	2016	23
Czechia	1993	2017	24	Czech Republic	1990	2016	27
Czechoslovakia	1968	1987	20				
Pakistan	1972	2017	31	Pakistan	1967	2016	50
East and West Pakistan	1962	1971	10				
Ethiopia	1995	2016	21	Ethiopia	2011	2016	6
Fmr Ethiopia	1962	1987	21				
Fmr Yugoslavia	1962	1987	26				
Germany	1991	2017	27	Germany	1970	2016	47
Fmr Dem. Rep. of Germany	1985	1987	3				
Fmr Fed. Rep. of Germany	1962	1990	29				
India	1975	2017	43	India	1960	2016	57
India, excl. Sikkim	1962	1974	13				
Panama	1978	2016	32	Panama	1960	2016	57
Fmr Panama, excl.Canal Zone	1962	1977	16				
Serbia	2005	2017	13	Serbia	1995	2016	22
Serbia and Montenegro	1992	2004	9				
State of Palestine	2007	2016	10				
				West Bank and Gaza	1994	2016	23
Sudan	2012	2015	2	Sudan	1960	2016	57
Fmr Sudan	1963	2011	37				
				South Sudan	2008	2015	8
Viet Nam	2000	2016	17	Vietnam	1986	2016	31
Fmr Rep. of Vietnam	1963	1973	11				
Yemen	2004	2015	12	Yemen, Rep.	1990	2016	27
Fmr Arab Rep. of Yemen	1975	1981	6				

Table C.7: Countries unmergeable due to name and time inconsistencies (economic datasets)

Countries coded as state actors in side A or B of the UCDP/PRIO Armed Conflict Dataset 18.1				System Membership Table (Table 3), UCDP/PRIO Armed Conflict Dataset Codebook p.15-20		
	year				year	
country	first	last	# obs.	State Name	first	last
Afghanistan	1978	2017	47	Afghanistan	1946	2012
Albania	1946	1946	2	Albania	1946	2012
Algeria	1963	2017	30	Algeria	1962	2012
Angola	1975	2017	36	Angola	1975	2012
Argentina	1955	1982	8	Argentina	1946	2012
				Armenia	1991	2012
Australia	2003	2003	2	Australia	1946	2012
				Austria	1946	2012
Azerbaijan	1991	2017	15	Azerbaijan	1991	2012
				Bahamas	1973	2012
				Bahrain	1971	2012
Bangladesh	1975	2017	21	Bangladesh	1971	2012
				Barbados	1966	2012
				Belarus (Byelorussia)	1991	2012
				Belgium	1946	2012
				Belize	1981	2012
				Benin	1960	2012
				Bhutan	1949	2012
Bolivia	1946	1967	3	Bolivia	1946	2012
Bosnia-Herzegovina	1992	1995	9	Bosnia-Herzegovina	1992	2012
				Botswana	1966	2012
				Brazil	1946	2012
				Brunei	1984	2012
				Bulgaria	1946	2012
Burkina Faso	1985	1987	3	Burkina Faso (Upper Volta)	1960	2012
Burundi	1965	2015	19	Burundi	1962	2012
Cambodia (Kampuchea)	1967	2011	42	Cambodia (Kampuchea)	1953	2012
Cameroon	1960	2017	10	Cameroon	1960	2012
				Canada	1946	2012
				Cape Verde	1975	2012
Central African Republic	2001	2013	8	Central African Republic	1960	2012
Chad	1966	2017	43	Chad	1960	2012
Chile	1973	1973	1	Chile	1946	2012
China	1946	2008	45	China	1946	2012
Colombia	1964	2016	53	Colombia	1946	2012
Comoros	1989	1997	2	Comoros	1975	2012
Congo	1993	2016	6	Congo	1960	2012

Table C.8: Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, A-Co

Countries coded as state actors in side A or B of the UCDP/PRIO Armed Conflict Dataset 18.1				System Membership Table (Table 3), UCDP/PRIO Armed Conflict Dataset Codebook p.15-20		
	year				year	
country	first	last	# obs.	State Name	first	last
DR Congo (Zaire)	1960	2017	30	Congo, Democratic Republic of (Zaire)	1960	2012
Costa Rica	1948	1948	1	Costa Rica	1946	2012
Ivory Coast	2002	2011	4	Cote D'Ivoire	1960	2012
Croatia	1992	1995	3	Croatia	1991	2012
Cuba	1953	1961	5	Cuba	1946	2012
Cyprus	1974	1974	2	Cyprus	1960	2012
				Czech Republic	1993	2012
				Czechoslovakia	1946	1992
				Denmark	1946	2012
Djibouti	1991	2008	7	Djibouti	1977	2012
Dominican Republic	1965	1965	1	Dominican Republic	1946	2012
				East Timor	2002	2012
Ecuador	1995	1995	2	Ecuador	1946	2012
Egypt	1948	2017	29	Egypt	1946	2012
El Salvador	1969	1991	16	El Salvador	1946	2012
				Equatorial Guinea	1968	2012
Eritrea	1997	2016	12	Eritrea	1993	2012
				Estonia	1991	2012
Ethiopia	1960	2016	131	Ethiopia	1946	2012
				Fiji	1970	2012
				Finland	1946	2012
France	1946	1962	55	France	1946	2012
Gabon	1964	1964	1	Gabon	1960	2012
Gambia	1981	1981	1	Gambia	1965	2012
Georgia	1991	2008	8	Georgia	1991	2012
				German Democratic Republic	1949	1990
				German Federal Republic	1949	2012
Ghana	1966	1983	3	Ghana	1957	2012
Greece	1946	1949	4	Greece	1946	2012
Grenada	1983	1983	2			
Guatemala	1949	1995	34	Guatemala	1946	2012
Guinea	2000	2001	2	Guinea	1958	2012
Guinea-Bissau	1998	1999	2	Guinea-Bissau	1974	2012
				Guyana	1966	2012
Haiti	1989	2004	3	Haiti	1946	2012
Honduras	1957	1969	3	Honduras	1946	2012
Hungary	1956	1956	2	Hungary	1946	2012
Hyderabad	1947	1948	4			
				Iceland	1946	2012
India	1948	2017	220	India	1947	2012
Indonesia	1950	2005	52	Indonesia	1946	2012
Iran	1946	2017	62	Iran (Persia)	1946	2012
Iraq	1948	2017	78	Iraq	1946	2012

Table C.9: Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, Co-Ira

Countries coded as state actors in side A or B of the UCDP/PRIO Armed Conflict Dataset 18.1				System Membership Table (Table 3), UCDP/PRIO Armed Conflict Dataset Codebook p.15-20		
	year				year	
country	first	last	# obs.	State Name	first	last
				Ireland	1946	2012
Israel	1948	2014	86	Israel	1948	2012
				Italy/Sardinia	1946	2012
				Jamaica	1962	2012
				Japan	1946	2012
Jordan	1948	2016	6	Jordan	1946	2012
				Kazakhstan	1991	2012
Kenya	1982	2017	4	Kenya	1963	2012
				Kosovo	2008	2012
Kuwait	1990	1991	2	Kuwait	1961	2012
				Kyrgyz Republic	1991	2012
Laos	1959	1990	22	Laos	1954	2012
				Latvia	1991	2012
Lebanon	1948	2017	17	Lebanon	1946	2012
Lesotho	1998	1998	1	Lesotho	1966	2012
Liberia	1980	2003	7	Liberia	1946	2012
Libya	1987	2017	8	Libya	1951	2012
				Lithuania	1991	2012
				Luxembourg	1946	2012
Macedonia, FYR	2001	2001	1	Macedonia (FRY)	1991	2012
Madagascar	1971	1971	1	Madagascar (Malagasy)	1960	2012
				Malawi	1964	2012
Malaysia	1958	2013	15	Malaysia	1957	2012
				Maldives	1965	2012
Mali	1985	2017	18	Mali	1960	2012
				Malta	1964	2012
Mauritania	1975	2011	6	Mauritania	1960	2012
				Mauritius	1968	2012
Mexico	1994	1996	2	Mexico	1946	2012
Moldova	1992	1992	1	Moldova	1991	2012
				Mongolia	1946	2012
				Montenegro	2006	2012
Morocco	1963	1989	17	Morocco	1956	2012
Mozambique	1977	2016	18	Mozambique	1975	2012
Myanmar (Burma)	1948	2017	275	Myanmar (Burma)	1948	2012
				Namibia	1990	2012
Nepal	1960	2006	14	Nepal	1946	2012
Netherlands	1946	1962	5	Netherlands	1946	2012
				New Zealand	1946	2012
Nicaragua	1957	1990	13	Nicaragua	1946	2012
Niger	1991	2017	10	Niger	1960	2012
Nigeria	1966	2017	20	Nigeria	1960	2012
North Korea	1949	1953	10	North Korea	1948	2012
				Norway	1946	2012
Oman	1957	1975	8	Oman	1946	2012

Table C.10: Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, Ire-O

Countries coded as state actors in side A or B of the UCDP/PRIO Armed Conflict Dataset 18.1				System Membership Table (Table 3), UCDP/PRIO Armed Conflict Dataset Codebook p.15-20		
	year				year	
country	first	last	# obs.	State Name	first	last
Pakistan	1948	2017	55	Pakistan	1947	2012
Panama	1989	1989	3	Panama	1946	2012
Papua New Guinea	1990	1996	6	Papua New Guinea	1975	2012
Paraguay	1947	1989	3	Paraguay	1946	2012
Peru	1965	2010	24	Peru	1946	2012
Philippines	1946	2017	104	Philippines	1946	2012
				Poland	1946	2012
Portugal	1961	1974	36	Portugal	1946	2012
				Qatar	1971	2012
Rumania	1989	1989	1	Rumania	1946	2012
Russia (Soviet Union)	1946	2017	44	Russia (Soviet Union)	1946	2012
Rwanda	1990	2016	17	Rwanda	1962	2012
Saudi Arabia	1979	1979	1	Saudi Arabia	1946	2012
Senegal	1990	2011	10	Senegal	1960	2012
Serbia (Yugoslavia)	1991	1999	5	Serbia	2006	2012
				Yugoslavia (Serbia)	1946	2006
Sierra Leone	1991	2001	11	Sierra Leone	1961	2012
				Singapore	1965	2012
				Slovakia	1993	2012
				Slovenia	1992	2012
				Solomon Islands	1978	2012
Somalia	1964	2017	32	Somalia	1960	2012
South Africa	1966	1988	30	South Africa	1946	2012
South Korea	1949	1953	5	South Korea	1948	2012
South Sudan	2011	2017	9	South Sudan	2011	2012
Spain	1957	1991	11	Spain	1946	2012
Sri Lanka	1971	2009	27	Sri Lanka	1948	2012
Sudan	1963	2017	49	Sudan	1956	2012
Suriname	1987	1987	1	Surinam	1975	2012
				Swaziland	1968	2012
				Sweden	1946	2012
				Switzerland	1946	2012
Syria	1948	2017	27	Syria	1946	2012
Taiwan	1949	1958	4	Taiwan	1949	2012
Tajikistan	1992	2011	10	Tajikistan	1991	2012
Tanzania	1978	1978	2	Tanzania/Tanganyika	1961	2012
Thailand	1946	2017	32	Thailand	1946	2012
				Tibet	1946	1950
Togo	1986	1986	1	Togo	1960	2012
Trinidad and Tobago	1990	1990	1	Trinidad and Tobago	1962	2012
Tunisia	1961	2016	3	Tunisia	1956	2012
Turkey	1974	2017	41	Turkey/Ottoman Empire	1946	2012
				Turkmenistan	1991	2012

Table C.11: Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, P-T

Countries coded as state actors in side A or B of the UCDP/PRIO Armed Conflict Dataset 18.1				System Membership Table (Table 3), UCDP/PRIO Armed Conflict Dataset Codebook p.15-20		
	year				year	
country	first	last	# obs.	State Name	first	last
Uganda	1971	2017	41	Uganda	1962	2012
Ukraine	2014	2017	7	Ukraine	1991	2012
				United Arab Emirates	1971	2012
United Kingdom	1946	2003	56	United Kingdom	1946	2012
United States of America	1950	2017	23	United States of America	1946	2012
Uruguay	1972	1972	1	Uruguay	1946	2012
Uzbekistan	1999	2004	3	Uzbekistan	1991	2012
Venezuela	1962	1992	3	Venezuela	1946	2012
Vietnam (North Vietnam)	1965	1988	24	Vietnam, Democratic Republic of	1954	2012
South Vietnam	1955	1975	32	Vietnam, Republic of	1954	1975
Yemen (North Yemen)	1948	2017	27	Yemen (Arab Republic of Yemen)	1946	2012
South Yemen	1972	1986	5	Yemen, People's Republic of	1967	1990
				Zambia	1964	2012
				Zanzibar	1963	1964
Zimbabwe (Rhodesia)	1967	1979	9	Zimbabwe (Rhodesia)	1965	2012

Table C.12: Comparison of country coding units in UCDP/PRIO Armed Conflict dataset 18.1 and the coding units supplied in the relevant code book, U-Z

Appendix D

Heterogeneity matters

D.1 List of countries and regions included in the analysis

region	country
Latin America	Argentina , Bolivia , Chile, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Suriname, Trinidad and Tobago, Uruguay
Middle East & North Africa	Algeria, Iran, Morocco, Turkey
Sub-Saharan Africa	Botswana, Burkina Faso, Burundi, Central African Republic, Chad, Democratic Republic of Congo, Gabon, Ghana, Ivory Coast, Kenya, Madagascar, Mauritania, Niger, Nigeria, Republic of the Congo, Rwanda, Senegal, South Africa, Sudan, Togo, Uganda
Western Europe, Australia & North America	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, United Kingdom, United States of America
Asia	China, India, Indonesia, Japan, Malaysia, Philippines, Singapore, South Korea, Sri Lanka, Thailand

Table D.1: Countries and regions included in the analysis

D.2 Regional Impulse Response Functions

The regional impulse response functions were computed as follows: In a first step the VAR-model is estimated separately for each country. This has the advantage of allowing for heterogeneous slope parameters across countries (as opposed to estimating a panel model with fixed effects). Secondly, for each country the given Cholesky-decomposition is employed and impulse response functions are calculated. Third, regional impulse response functions are derived by taking the average of the corresponding countries' point estimates. Confidence bands, however, may not be pooled accross countries by using averages since they are based on non-linear quantiles. Averages over confidence bands would yield unclear coverage, i.e. it would be unclear whether the confidence bands coverage retains its nominal size of 95%. Regional confidence bands are therefore computed employing the following variance decomposition:

For each forecasting step ($s = 1, 2, ..., 10$) the regional variance that consists of within and between variance:

$$\sigma^2_{s,regional} = \sigma^2_{s,within} + \sigma^2_{s,between}, \tag{D.1}$$

with

$$\sigma^2_{s,within} = \frac{1}{N}\sum_{i=1}^{N} variance_{i,s},$$

where $i = 1, 2, ..., N$ is the number of countries in the regional group and $variance_{i,s} = standarderror^2_{i,s} \cdot \tau$. τ refers to the degrees of freedom adjusted lengths of each country's timeseries, i.e. $\tau = T - L - L * K - 1$, where $T = 57$ years, lag length $L = 4$, number of control variables $K = 4$, and -1 for the intercept. Therefore, $\tau = 57 - 4 - 4 * 4 - 1 = 36$.

For the regional between variance the heterogeneity of the point estimates is exploited. It is computed using the folllowing formula:

$$\sigma^2_{s,between} = \frac{1}{N}\sum_{i=1}^{N}[IRF_{i,s} - \frac{1}{N}\sum_{i=1}^{N} IRF_{i,s}]^2$$

where $IRF_{i,s}$ are the point estimates obtained for each step $s = 1, 2, ..., 10$ and each country i in the region (N =total number of countries in the region).

Once $\sigma^2_{s,regional}$ is computed regional confidence bands for each step can be calculated as upper (lower) 95%: $\overline{IRF_s} + (-)\ \sigma^2_{s,regional} \cdot 1,96$, where $\overline{IRF_s}$ is the simple regional average of the point estimates.

Since the heterogeneity of responses is substantial across countries the regionally aggregated impulse response functions in Figure D.1 display no effect. Note that this result is solely driven by the regional heterogeneity of effects, not by their inexistence.

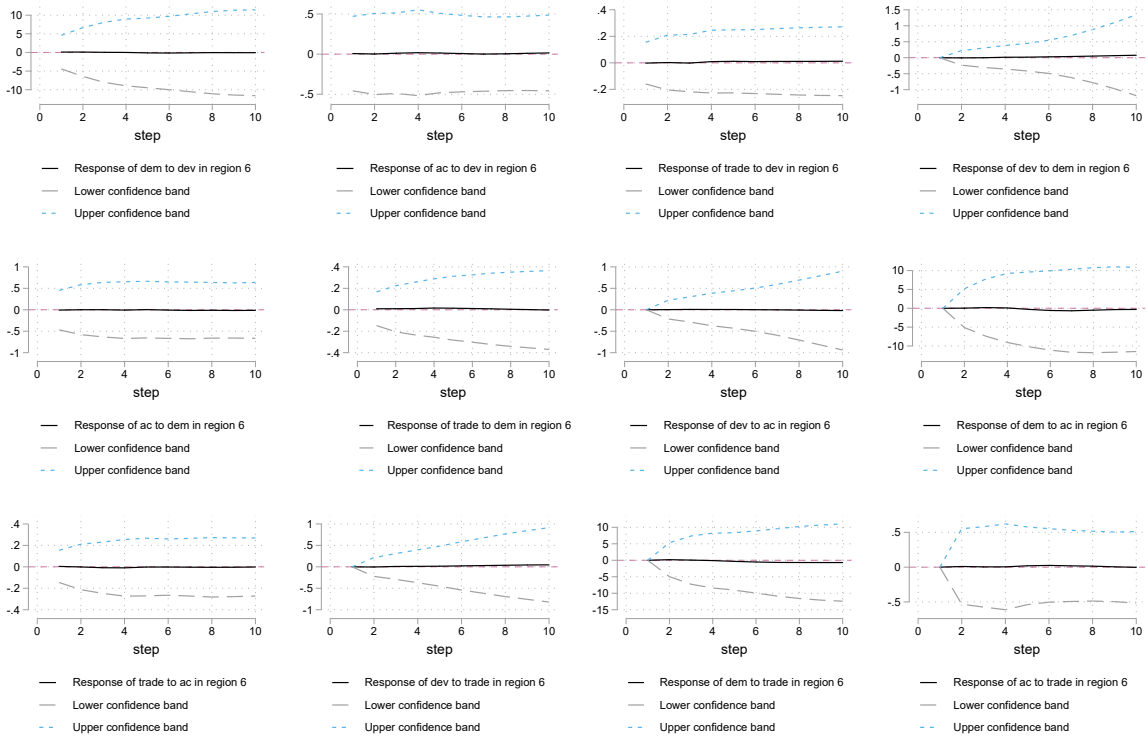


Figure D.1: Regional impulse response functions for Asia. "ac" stands for "armed conflict", "dev" for development, and "dem" for democracy

D.3 Impulse response functions - Baseline model

This section displays and discusses the impulse response function of the baseline model: VAR(4), ordering: development, democracy, conflict and trade.

For the sake of space only impulse response functions for countries with significant effects are shown. The impulse response functions are grouped by bilateral relationship.

Trade openness and development

Trade openness seems to have a negative effect on development only in a handful of countries. However, in these countries the effect is strong and persistent. Affected countries are exclusively African or Latin American. Nevertheless, the largest positive impacts of trade openness on development can as well be seen in Africa: the biggest effect is displayed by Ivory Coast. Smaller, but significant positive results are seen for Asian countries. Interestingly, the largest advocate for free trade profits from it: the US show a significant development effect from a positive shock in trade openness.

Development only has a negative result for three countries: Central African Republic, Nicaragua and Uruguay, with the biggest effect for Nicaragua. In turn, development has a positive effect on trade openness for a huge number of countries: across all regions countries trade openness is affected positively by a positive shock in development, with the largest impacts in Rwanda and Suriname, followed by the US, Algeria, Republic of the Congo and Senegal.

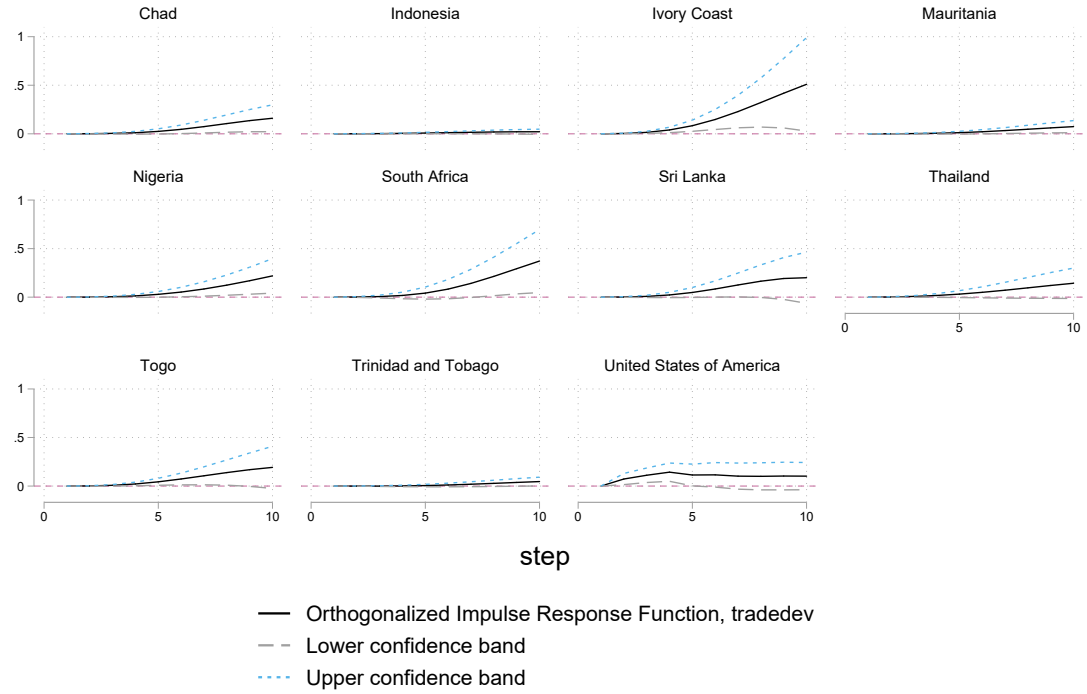


Figure D.2: Impulse response functions for countries with significant positive effects of trade on development during at least one step.

Summary: Development causes rise in trade openness in a large number of countries, while negative effects are rare. Trade openness does not necessarily cause rise in development: positive effects are rarely large. However, the same applies to negative effects of trade openness.

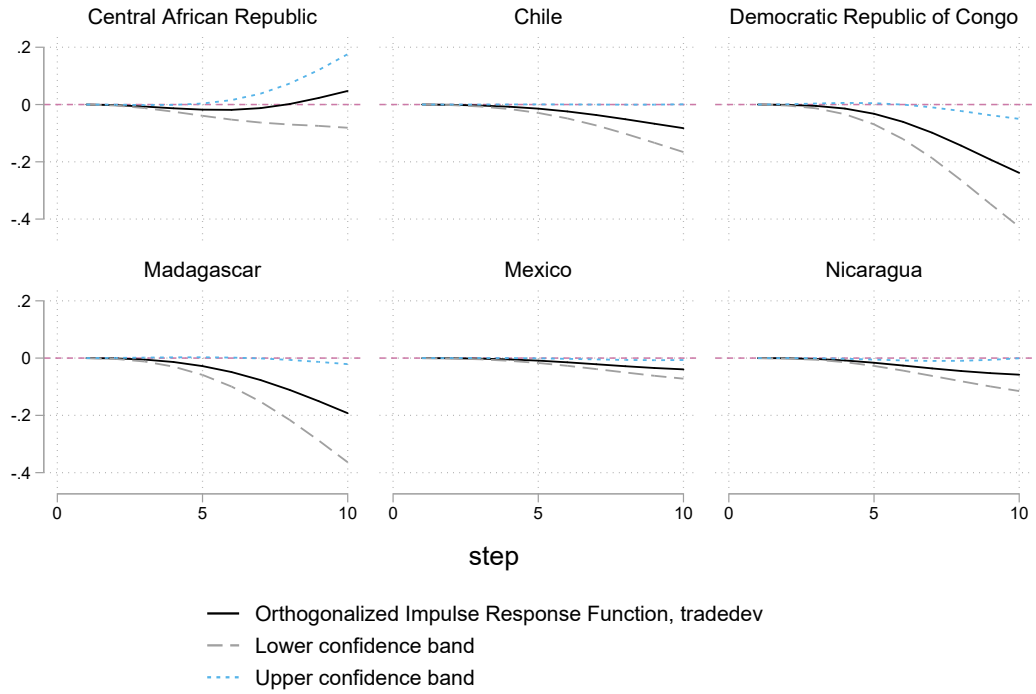


Figure D.3: Impulse response functions for countries with significant negative effects of trade on development during at least one step.

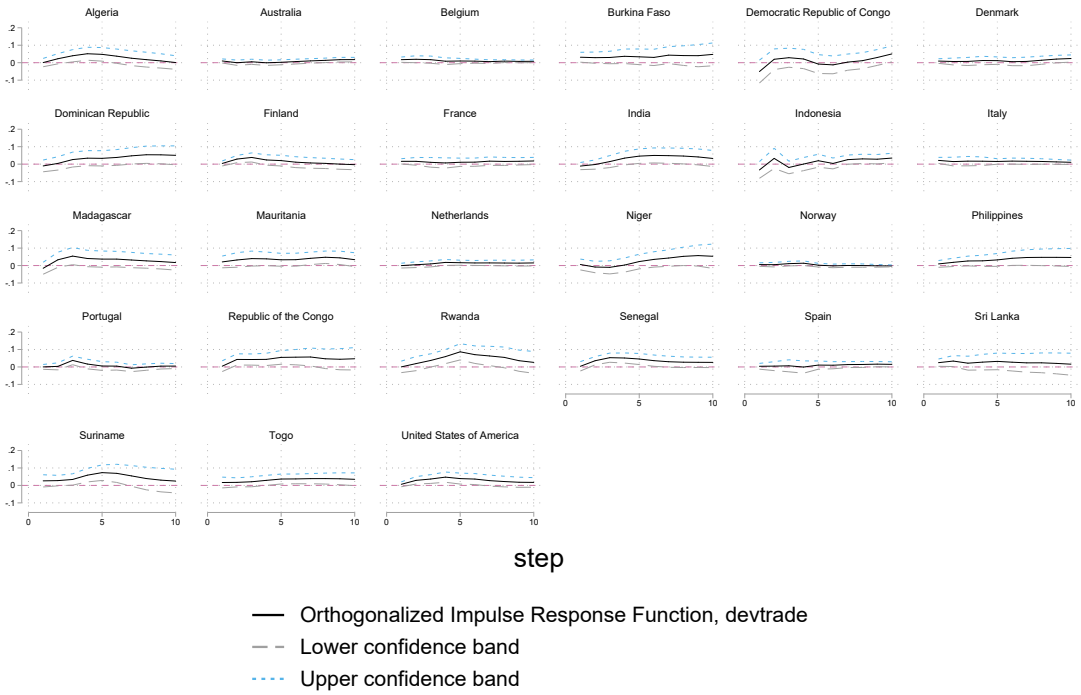


Figure D.4: Impulse response functions for countries with significant positive effects of development on trade during at least one step.

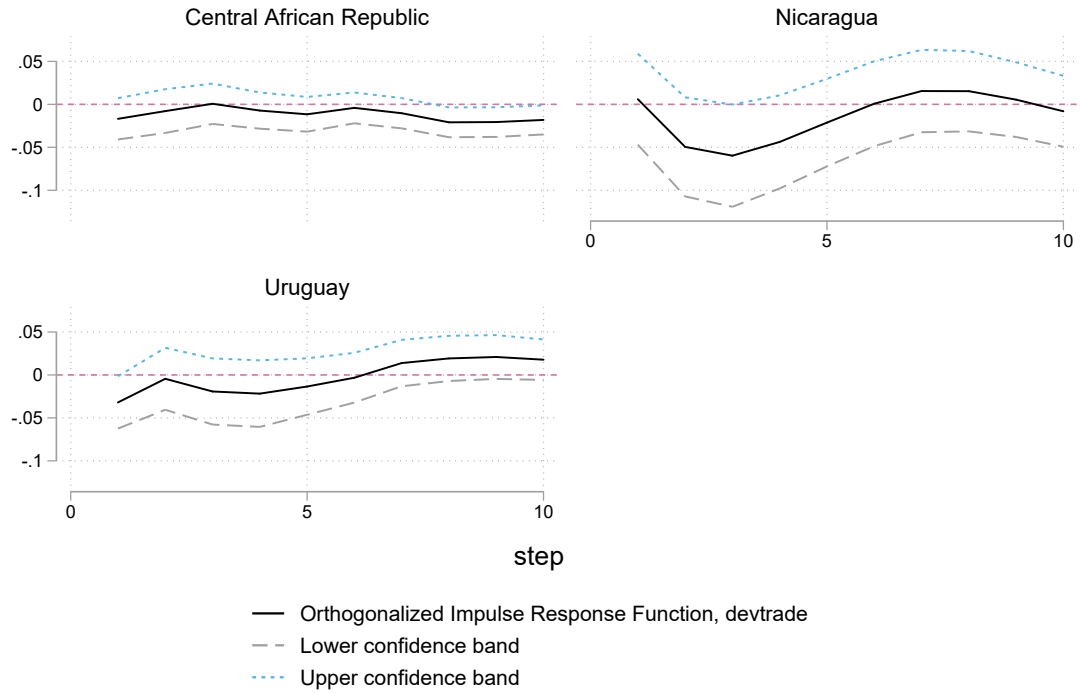


Figure D.5: Impulse response functions for countries with significant negative effects of development on trade during at least one step.

Trade openness and democracy

The highest positive effects of a shock in trade openness on democracy can be seen in Chile and Turkey. Especially Latin American countries display positive responses. While only 9 countries show positive results, 19 countries show negative responses in democracy after a positive shock in trade openness. This group is very heterogeneous and represents all 5 regions. Negative results are largest for Republic of Congo, South Korea and Niger.

Compared to the results seen before, responses for each direction are more balanced in this group. Trade openness is negatively affected by a shock in democracy in 15 countries of all regions, most notably Rwanda and Iran. Interestingly, especially China's trade openness profits from positive shocks in democracy. Other substantial effects of democracy on trade can be seen for Suriname, Republic of the Congo and Democratic Republic of the Congo.

Summary: The responses to a shock are heterogeneous in both parts of the bilateral relationship between trade openness and democracy.

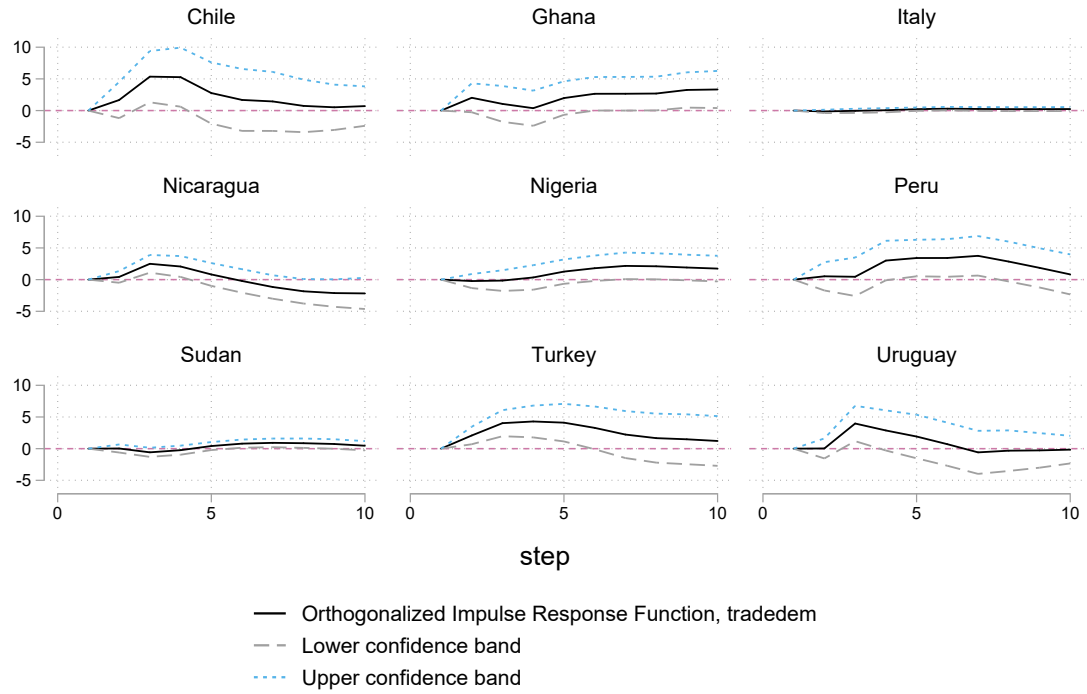


Figure D.6: Impulse response functions for countries with significant positive effects of trade on democracy during at least one step.

Trade openness and conflict

Overall a shock in trade openness leads to very heterogeneous responses across all countries. Interestingly, for this shock regional homogeneity is present to some extent. Four western states out of 13 countries show a positive response in armed conflict to a shock in trade. In Latin America, the only significant effects are negative deviations (of substantial magnitude, e.g. Nicaragua -.2) from the country means in 3 of 10 countries. All other regions contain countries with heterogeneous responses, which suggest conflict location as a potentially important variable to include in further research.

Overall, as well as in the reverse case, the responses of trade openness to a shock in armed conflict are heterogeneous across countries and regions. From the MENA region, only one country displays a significant response, which is Turkey, and the response is positive. However, no general result for this relationship can be stated.

Summary: Trade and conflict simultaneously affect each other. The effects are heterogeneous across countries, although there is some slight regional homogeneity.

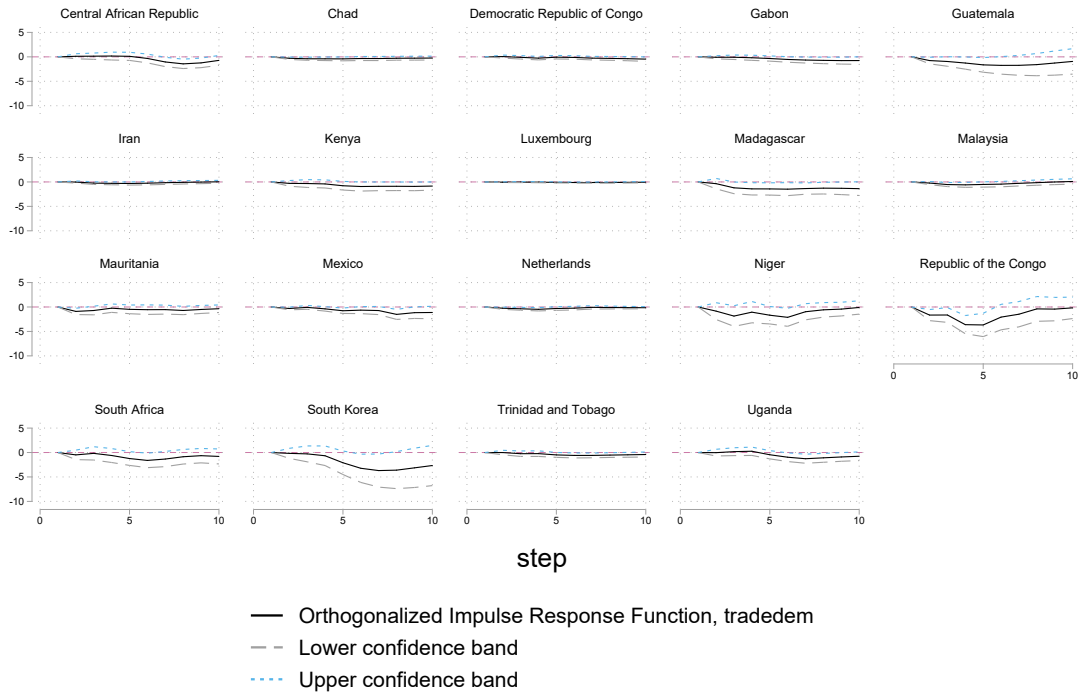


Figure D.7: Impulse response functions for countries with significant negative effects of trade on democracy during at least one step.

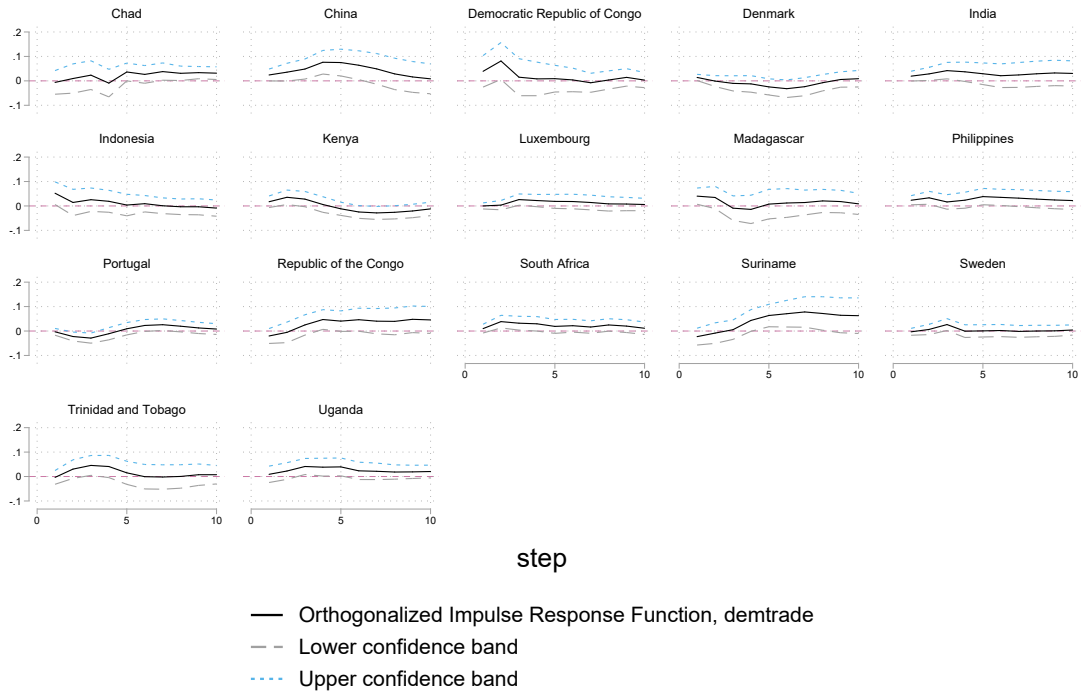


Figure D.8: Impulse response functions for countries with significant positive effects of democracy on trade during at least one step.

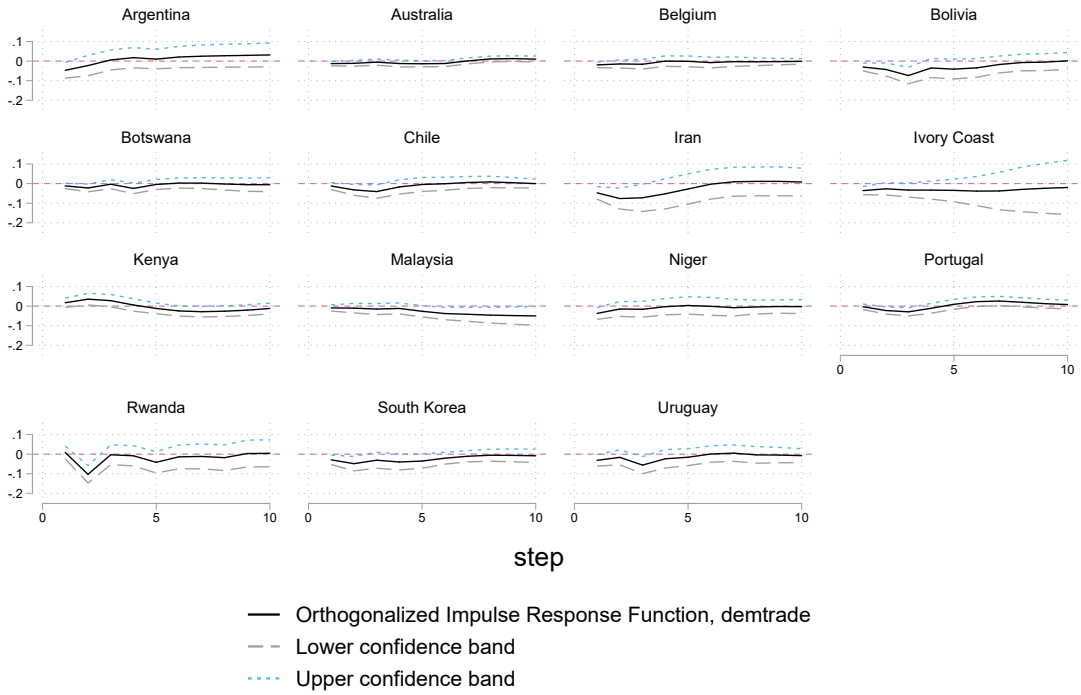


Figure D.9: Impulse response functions for countries with significant negative effects of democracy on trade during at least one step.

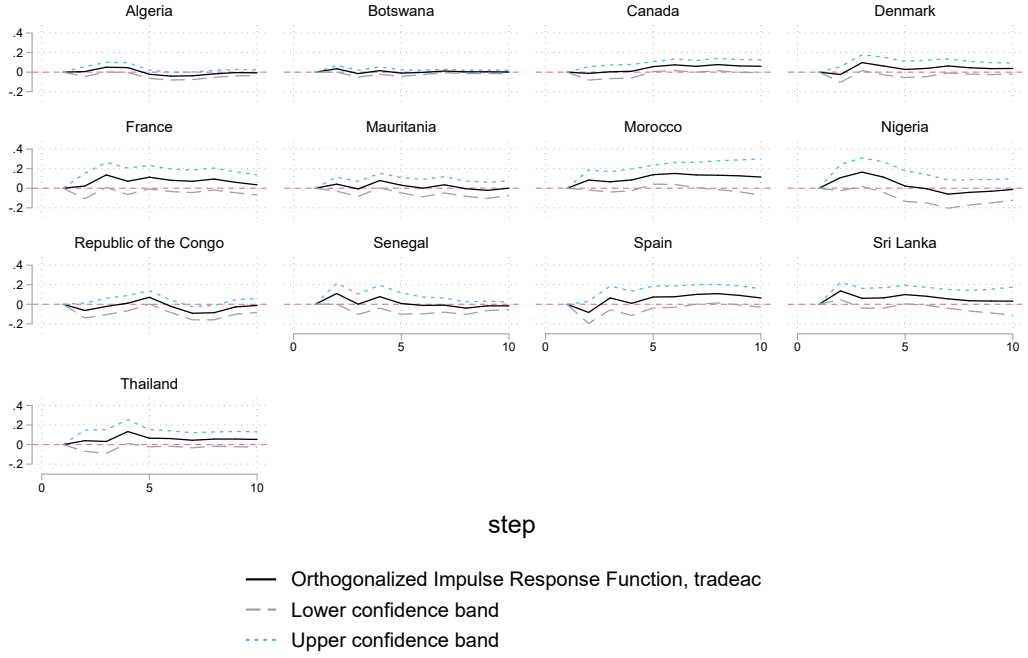


Figure D.10: Impulse response functions for countries with significant positive effects of trade on conflict during at least one step.

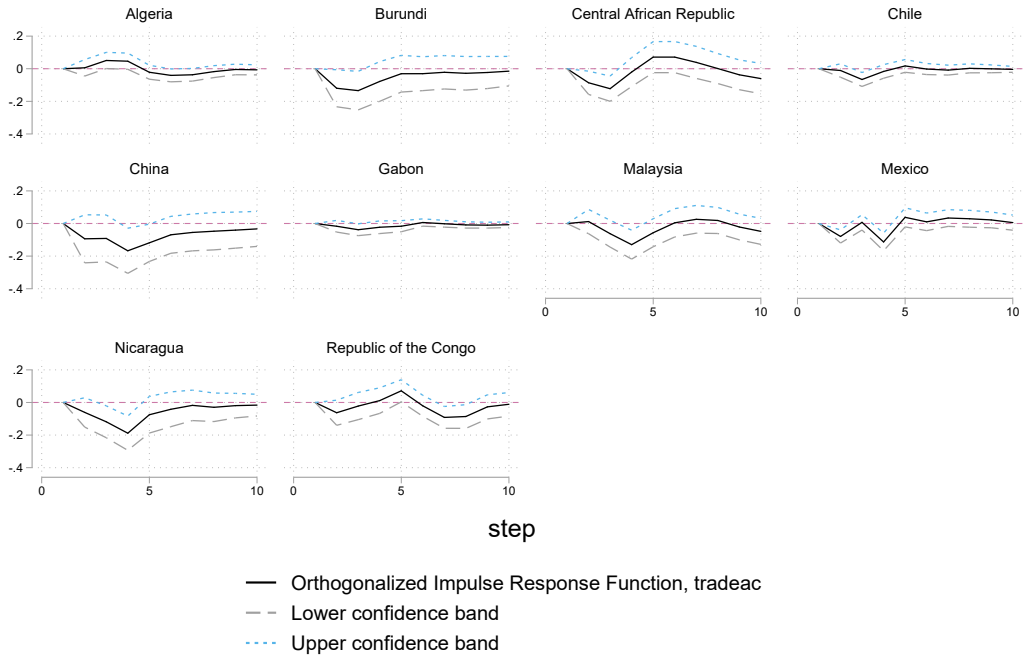


Figure D.11: Impulse response functions for countries with significant negative effects of trade on conflict during at least one step.

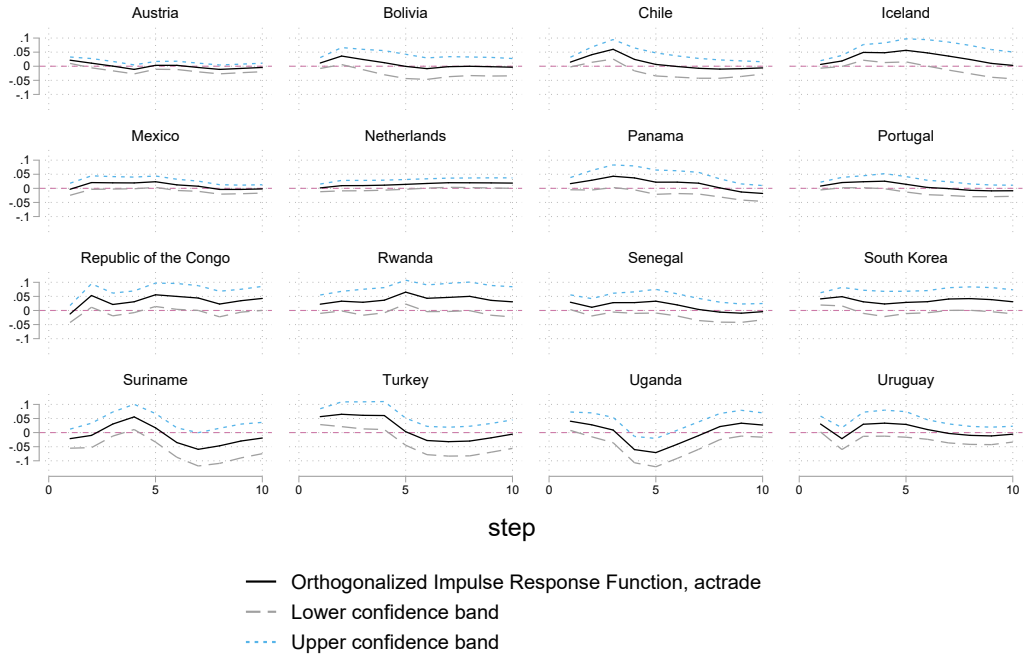


Figure D.12: Impulse response functions for countries with significant positive effects of conflict on trade during at least one step.

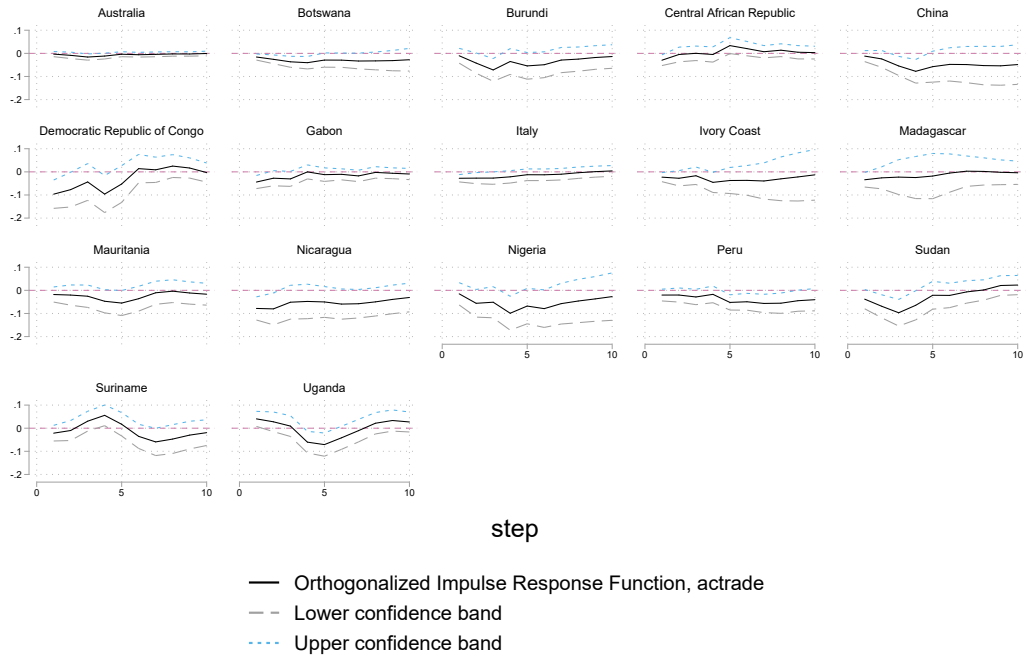


Figure D.13: Impulse response functions for countries with significant negative effects of conflict on trade during at least one step.

Democracy and development

In line with the slow moving and slow reacting nature of the two variables, democracy and development, observed effects of democracy on development mostly appear after 5 periods (i.e. in the long run) and they are long lasting (e.g. Ghana, Gabon, CAR, Senegal), that is they last for several periods in a row. In Sub-Saharan Africa, 5 of 21 countries show significant responses and all of them are positive. In Asia, only Sri Lanka shows an effect (positive between periods 4-9) In Western States 2 of 19 countries show a positive effect. Here, the effect is short (one or at most two periods) and occurs right after the shock (or within the first 5 periods). No effect is visible in any MENA country. On average democracy has mostly positive effects on development in this sample, if any. Over all regions the only countries which show a negative effect are three countries in Latin America. (Chile, Honduras and Mexico). These display long lasting negative effects of democracy on development.

The IRFs mostly provide evidence for positive effects of development on the level of democracy. In Sub-Saharan Africa and Western States the significant effects were entirely positive. However, there are two notable negative responses: in Indonesia a shock in development leads to a very strong negative response (up to -3 deviation from its country mean and over the first 7 periods). In Argentina, democracy also responds negatively (with deviations from its democracy average of up to -4.5 points) during periods 2 to 6. Morocco is an interesting case as there, democracy first responds negatively to a shock in development, but then adjusts and responds positively in periods 7-9. It is worth noting, though, that Morocco's response is rather small in magnitude, i.e. its democracy value never deviates more than +/-.3 from its average.

Summary: In total the impulse response functions point towards mutual positive effects of democracy on development and vice versa. There are, however, a few notable exceptions where

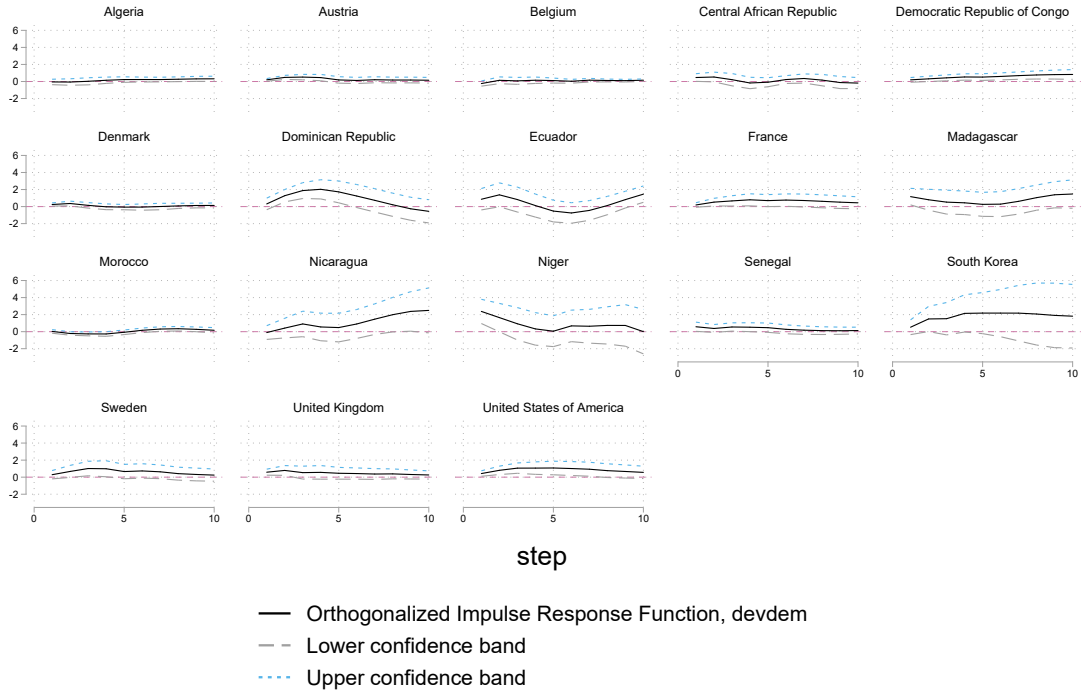


Figure D.14: Impulse response functions for countries with significant positive effects of development on democracy during at least one step.

negative responses occur. Due to the magnitude of these responses they should not be overlooked.

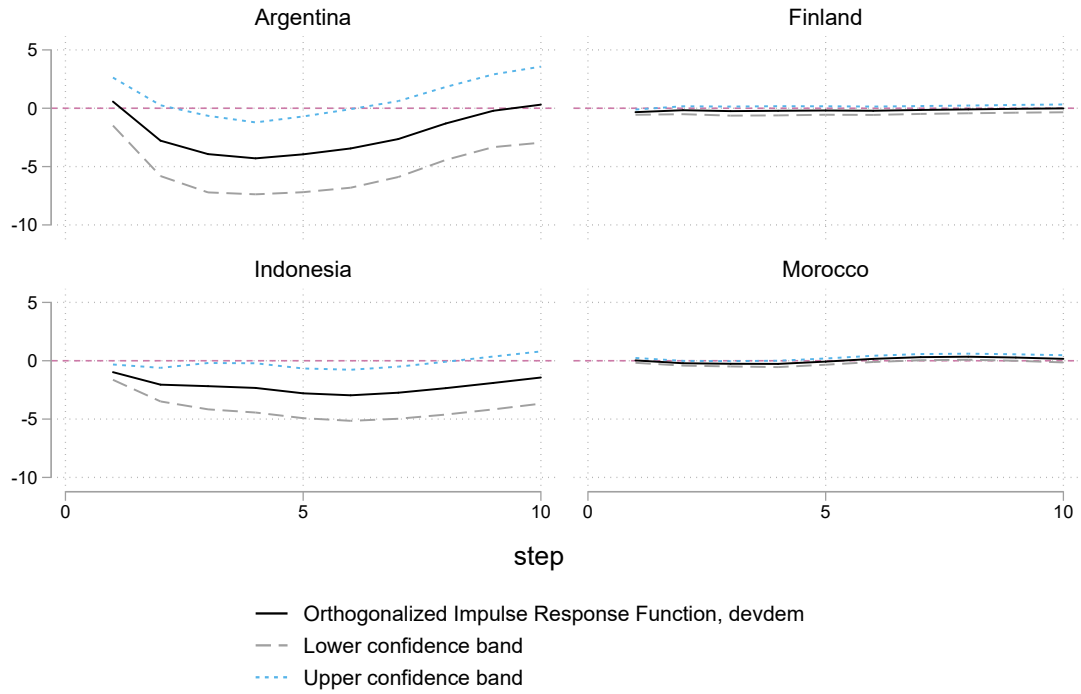


Figure D.15: Impulse response functions for countries with significant negative effects of development on democracy during at least one step.

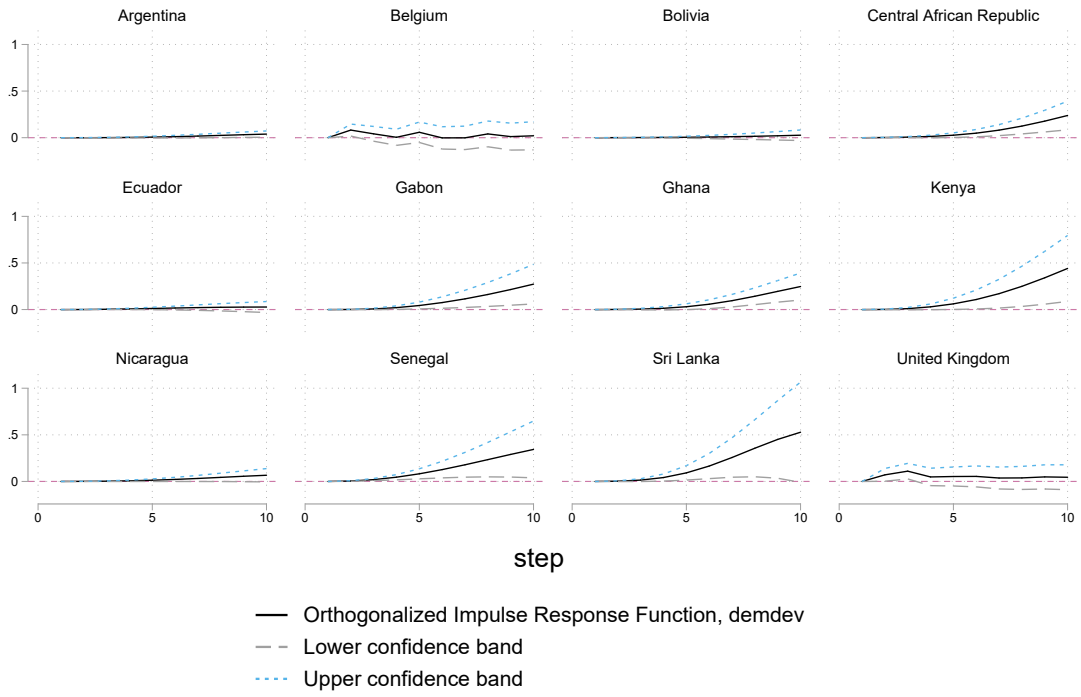


Figure D.16: Impulse response functions for countries with significant positive effects of democracy on development during at least one step.

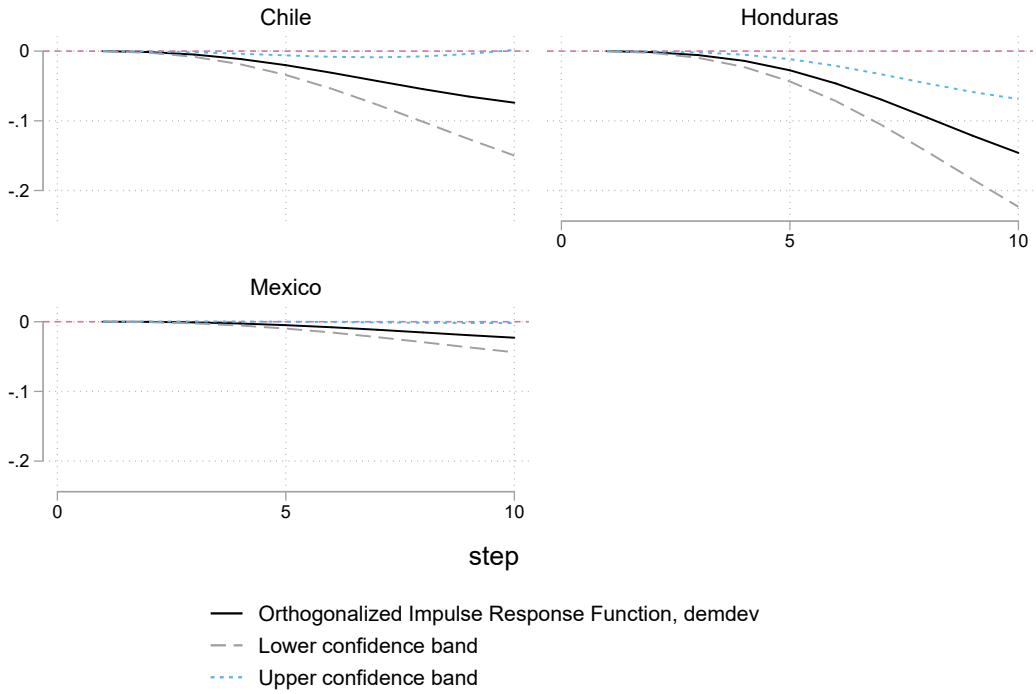


Figure D.17: Impulse response functions for countries with significant negative effects of democracy on development during at least one step.

Conflict and development

There are comparatively few countries in which a shock of armed conflict has significant effects on development. None of the countries in the MENA region, for example, show any significant effect. When there are effects they are mostly positive and appear only after 5 periods (for example in Nicaragua, Peru and Uruguay). In Sub-Saharan Africa this pattern is particularly pronounced for the Republic of Congo and Rwanda. In the Central African Republic on the other hand, conflict has a strong negative impact over the last three periods. Interestingly, the only country with a similar pattern is Austria. The Netherlands are a notable case as a shock in conflict there leads to up to +.18 positive deviation from its country average in development. In this model, both armed conflict and democracy have mostly positive effects on development. The magnitude of these effects is higher after shocks in democracy than in armed conflict.

Development has comparatively smaller effects on armed conflict than democracy, i.e. the responses are on average of a smaller magnitude (most of the countries deviate less than .1 points from their conflict average over the entire 10 periods) and they occur in a smaller number of countries. In several countries a positive shock in development leads to a positive response of armed conflict. In Western states this effect is very short lived and only visible in the first period. In Sub-Saharan Africa, Rwanda, the Democratic Republic of Congo and Sudan are most notable examples: their conflict values deviate around +0.12 over at least two periods after a positive shock in development. There are 7 countries for which a positive shock in development leads to a negative deviation from their country means in armed conflict. Once more conflict location suggests itself as an important factor to include in further research especially regarding the effects of conflict.

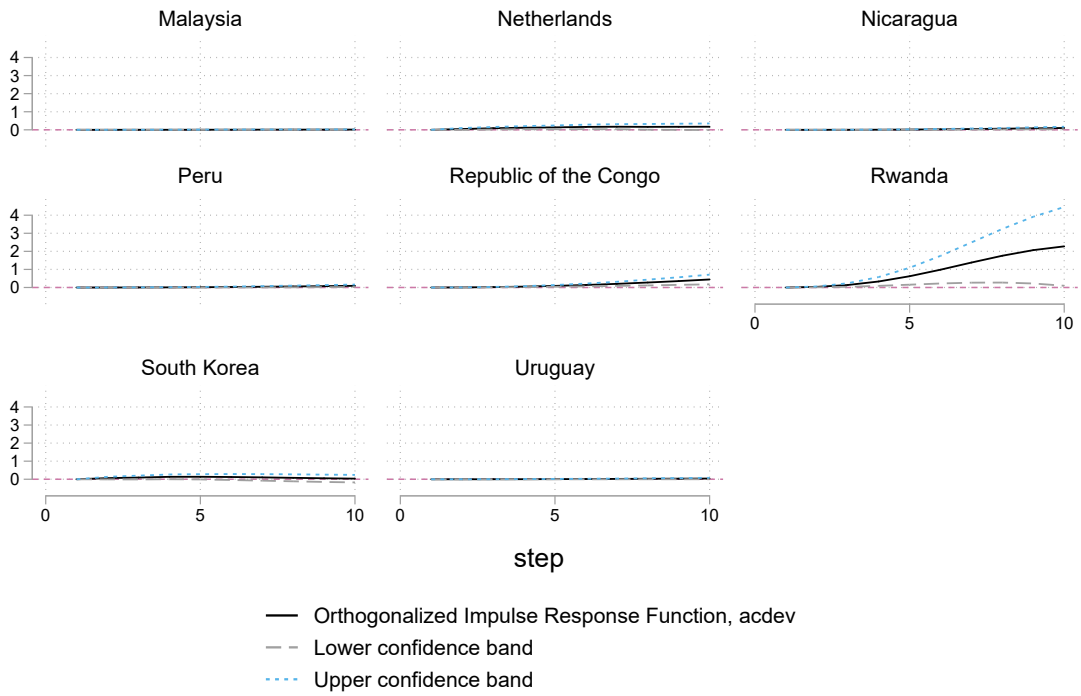


Figure D.18: Impulse response functions for countries with significant positive effects of conflict on development during at least one step.

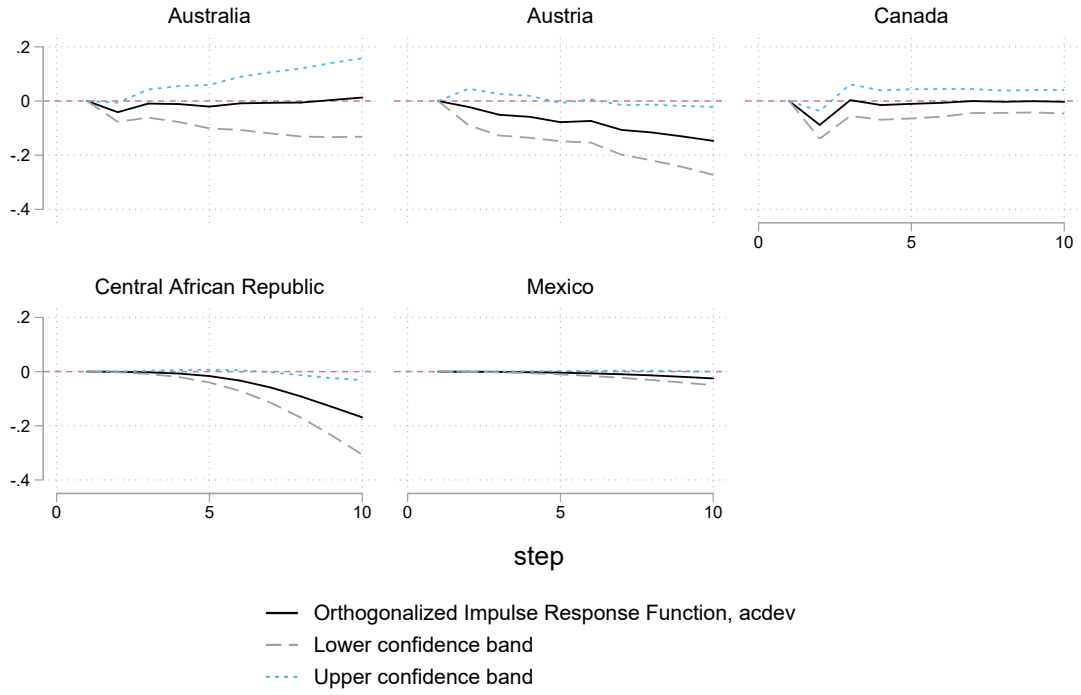


Figure D.19: Impulse response functions for countries with significant negative effects of conflict on development during at least one step.

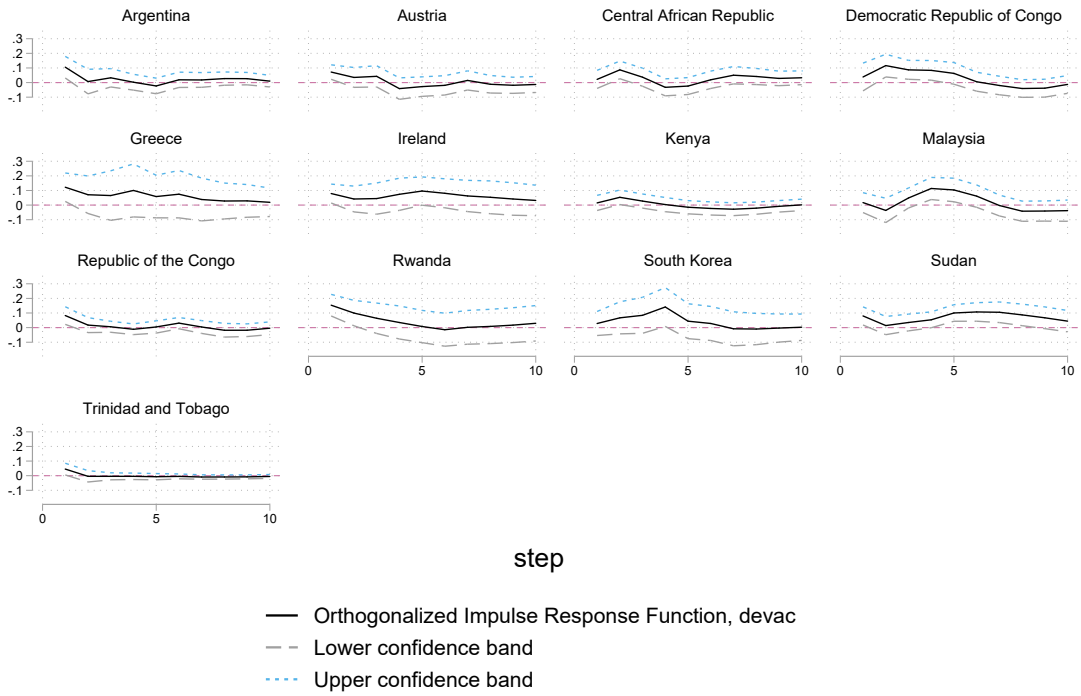


Figure D.20: Impulse response functions for countries with significant positive effects of development on conflict during at least one step.

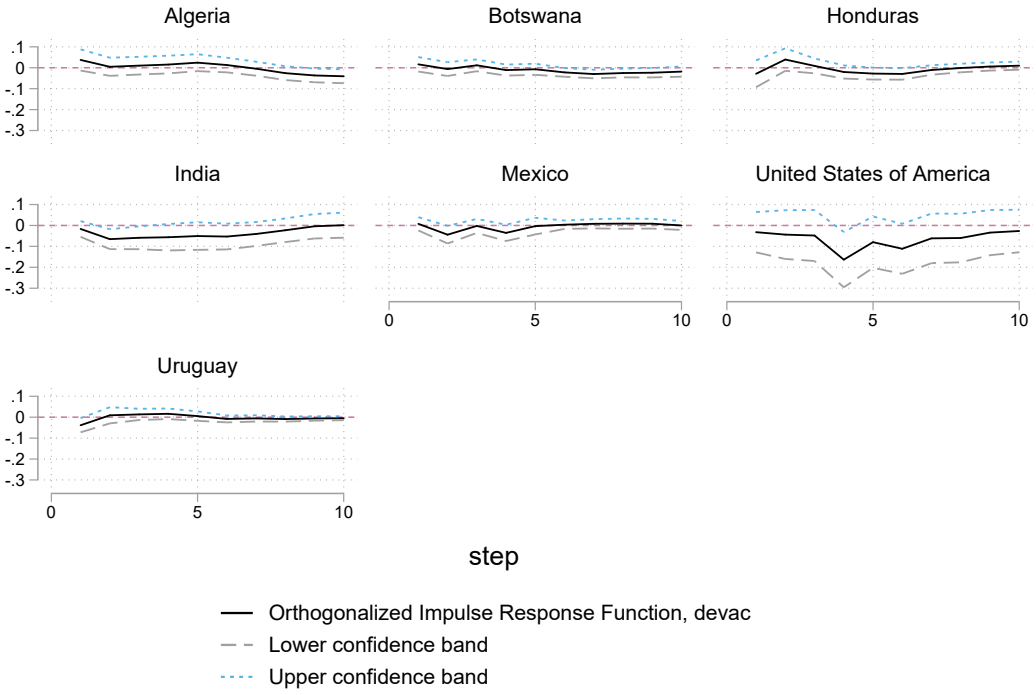


Figure D.21: Impulse response functions for countries with significant negative effects of development on conflict during at least one step.

Democracy and conflict

The effects of armed conflict on democracy are very heterogeneous across countries, in terms of magnitude as well as timing. There are positive and negative responses across all regions. Significant responses occur at all steps. There is no striking similarity or difference. The positive impact of armed conflict on democratization is especially striking in Central America: Mexico, Nicaragua and Panama display some of the largest significant results in the whole shock group, while results for regions 4, 5, and 6 show mixed results. In region 2, Turkey is the only country showing a significant result, which as well as in Central America shows that armed conflict increases democratization. Overall, timing of these shocks is very diverse.

The effects of a shock in democracy on conflict are highly heterogeneous across countries. For example, there is a significant negative effect at some point in the first five years in all MENA countries, particularly strong in Iran A strong significant positive effect on the other hand is visible in India and Malaysia within the first five periods. These are the only countries in Asia displaying any response. In general most reactions seem to happen in the first five periods. Responses after that mostly occur in Western States (Australia) or are very small (eg Togo, or Suriname)
Although some countries in region 5 show positive results, these results are very small (Norway, Netherlands, Spain, UK). A striking but as well intuitive result is that the biggest negative responses of armed conflict to a positive shock in democratization are seen in the western countries (Australia, Canada, Portugal): conflict involvement decreases the more democratic a country becomes.
The reverse is true for regions 4 and 6, where a positive shock in democratization leads to more Armed Conflict.

Summary: The impulse response functions show evidence for both armed conflict affecting democracy levels and vice -versa. In general the responses of democracy occur over longer periods of time whereas conflict reacts rather punctually (few steps at a time). There is evidence of simultaneity in a wide range of countries (e.g. Morocco: democracy has a negative impact on armed conflict and armed conflict has a negative impact on democracy; or India and Malaysia where during the first five steps democracy has a positive effect on armed conflict and so does armed conflict on democracy). This simultaneity can at times take counterintuitive forms, e.g. in South Africa: armed conflict has a negative effect on democracy in the long run. In the short run democracy has a positive effect on armed conflict.

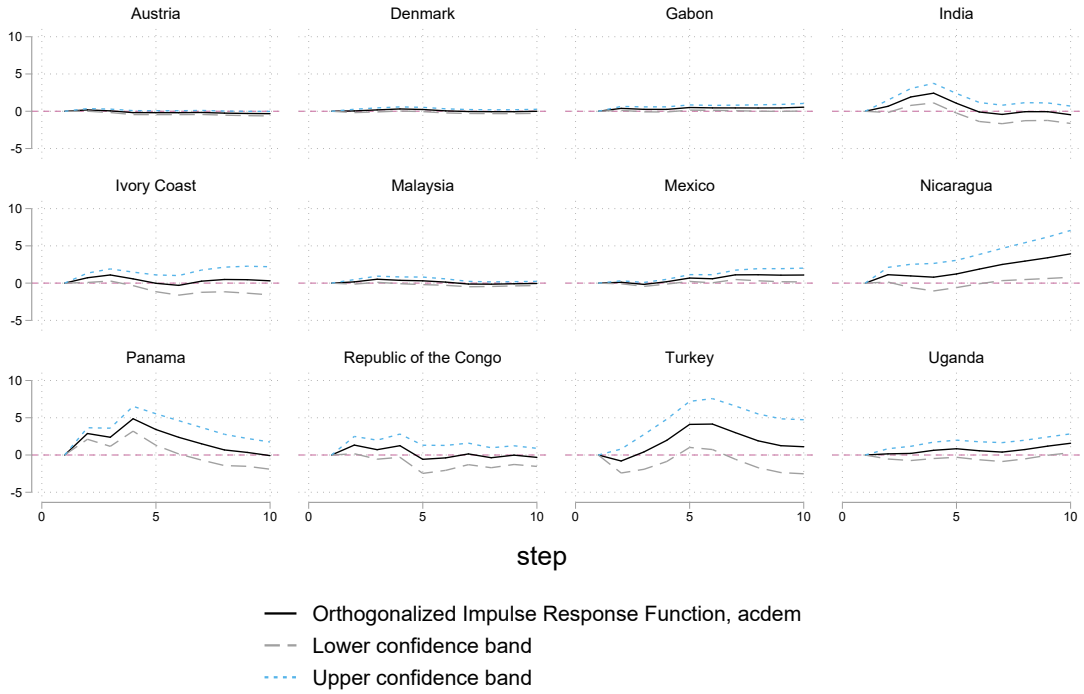


Figure D.22: Impulse response functions for countries with significant positive effects of conflict on democracy during at least one step.

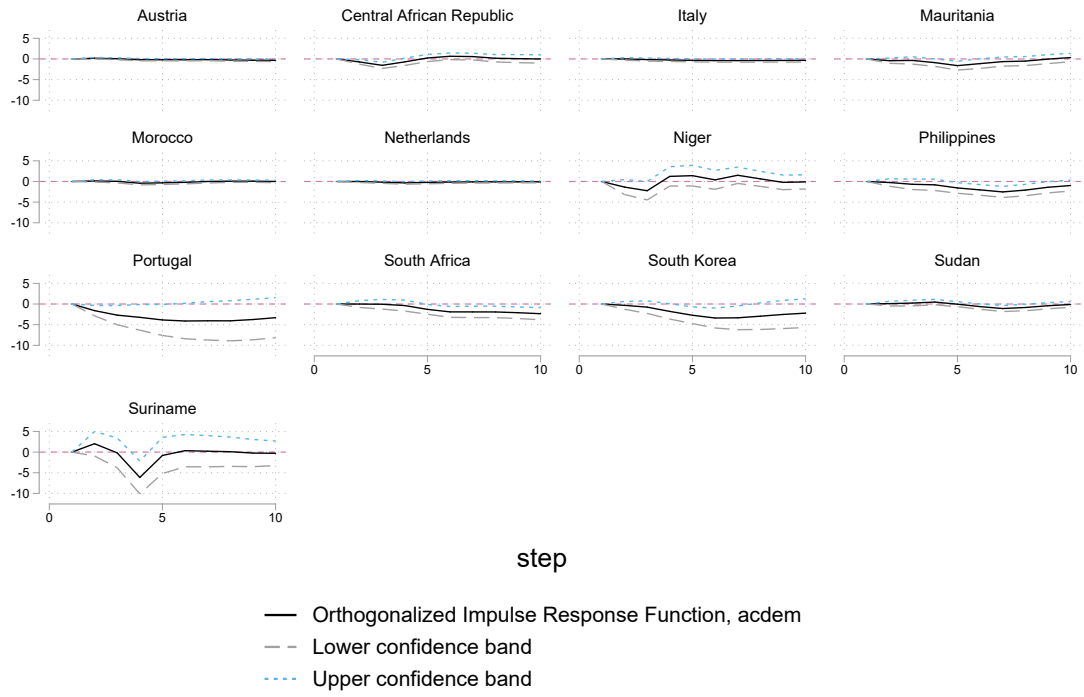


Figure D.23: Impulse response functions for countries with significant negative effects of conflict on democracy during at least one step.

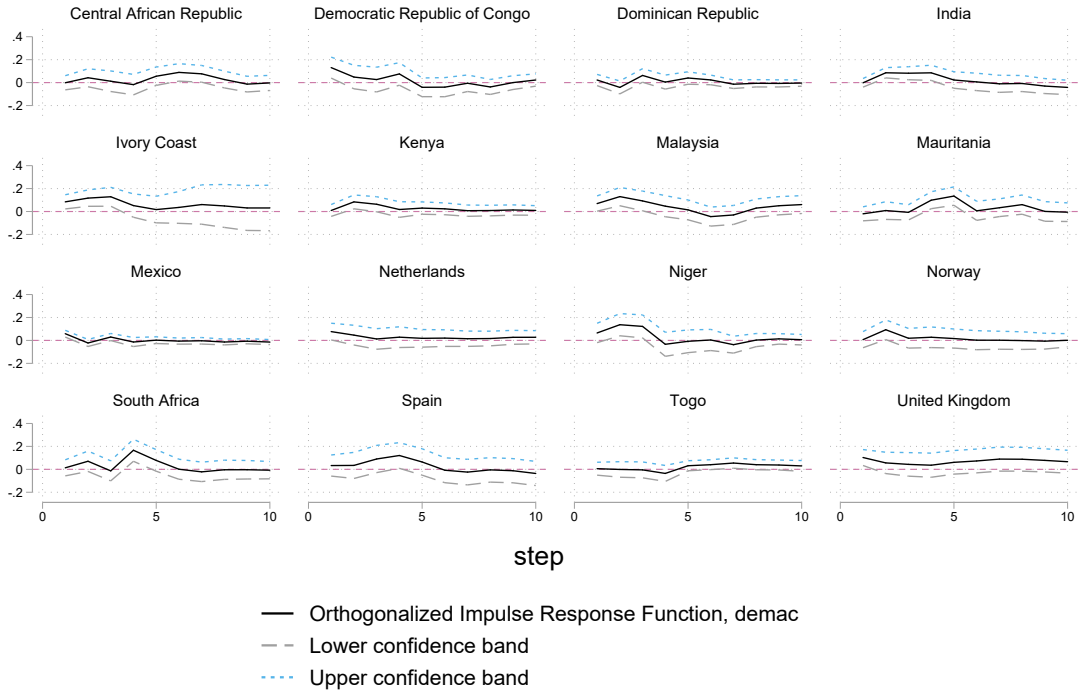


Figure D.24: Impulse response functions for countries with significant positive effects of democracy on conflict during at least one step.

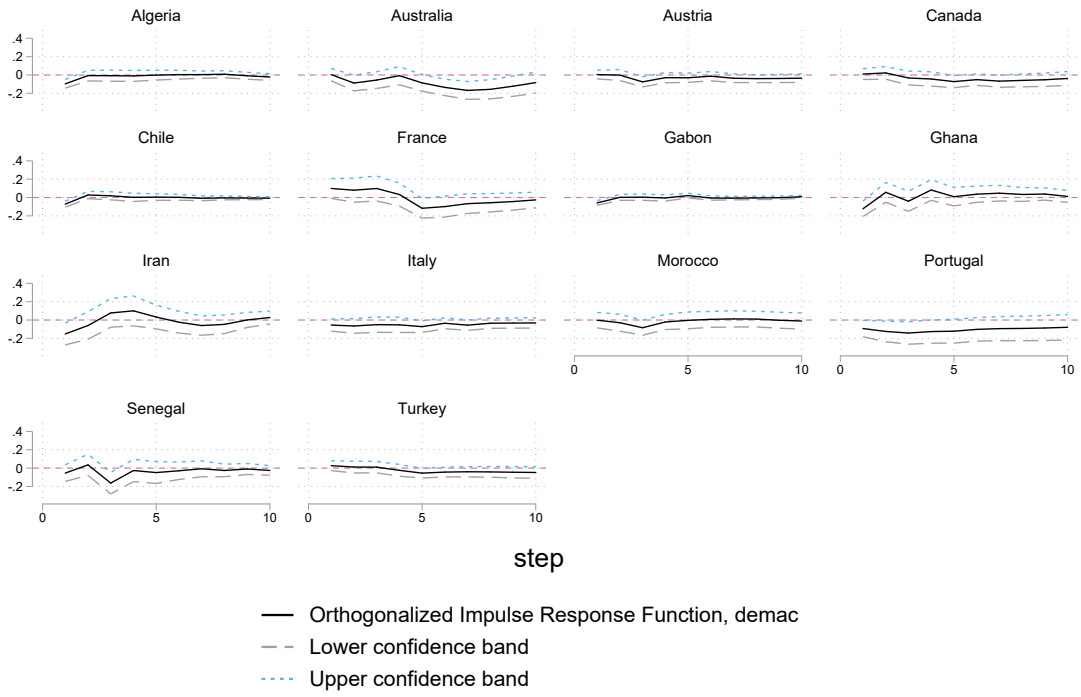


Figure D.25: Impulse response functions for countries with significant negative effects of democracy on conflict during at least one step.

D.4 Robustness Checks

On the following pages tables with the number of countries responding with positive, negative or no changes to a given shock are presented for all model specification discussed in Section 5.6:

1. VAR(4) model in first difference
2. VAR(4) model with alternative ordering (democracy, development, trade and conflict)
3. VAR(4) model with GDP per capita growth instead of female life expectancy
4. VAR(4) model with logged GDP per capita instead of female life expectancy
5. VAR(8) model
6. VAR(10) model

Robustness Check: Model in first differences, impulse response functions are accumulated IRFs, ordering: development, democracy, conflict and trade

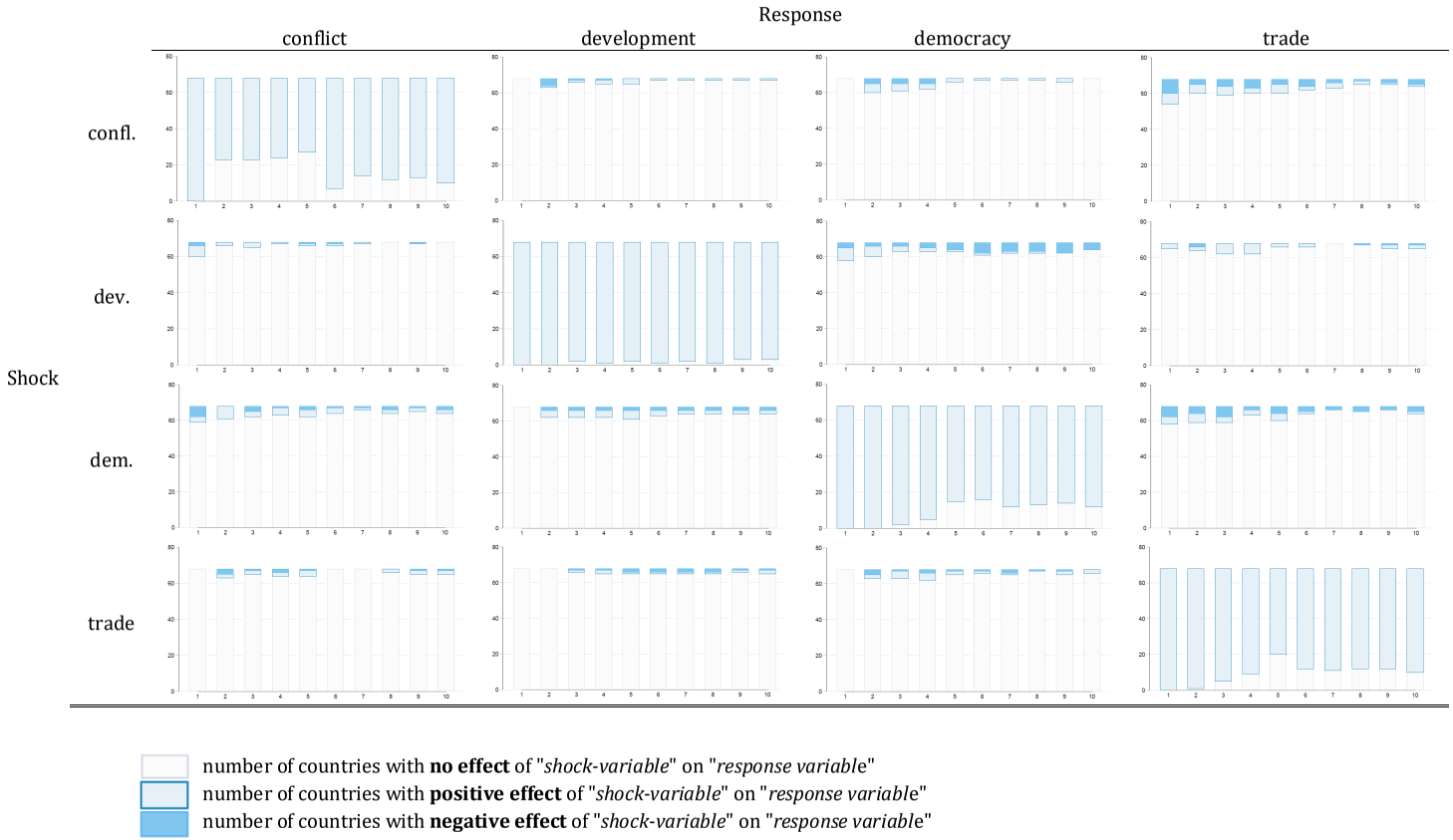


Figure D.26: Robustness Check - Model in first differences: Number of countries with non/significant responses to the given shocks

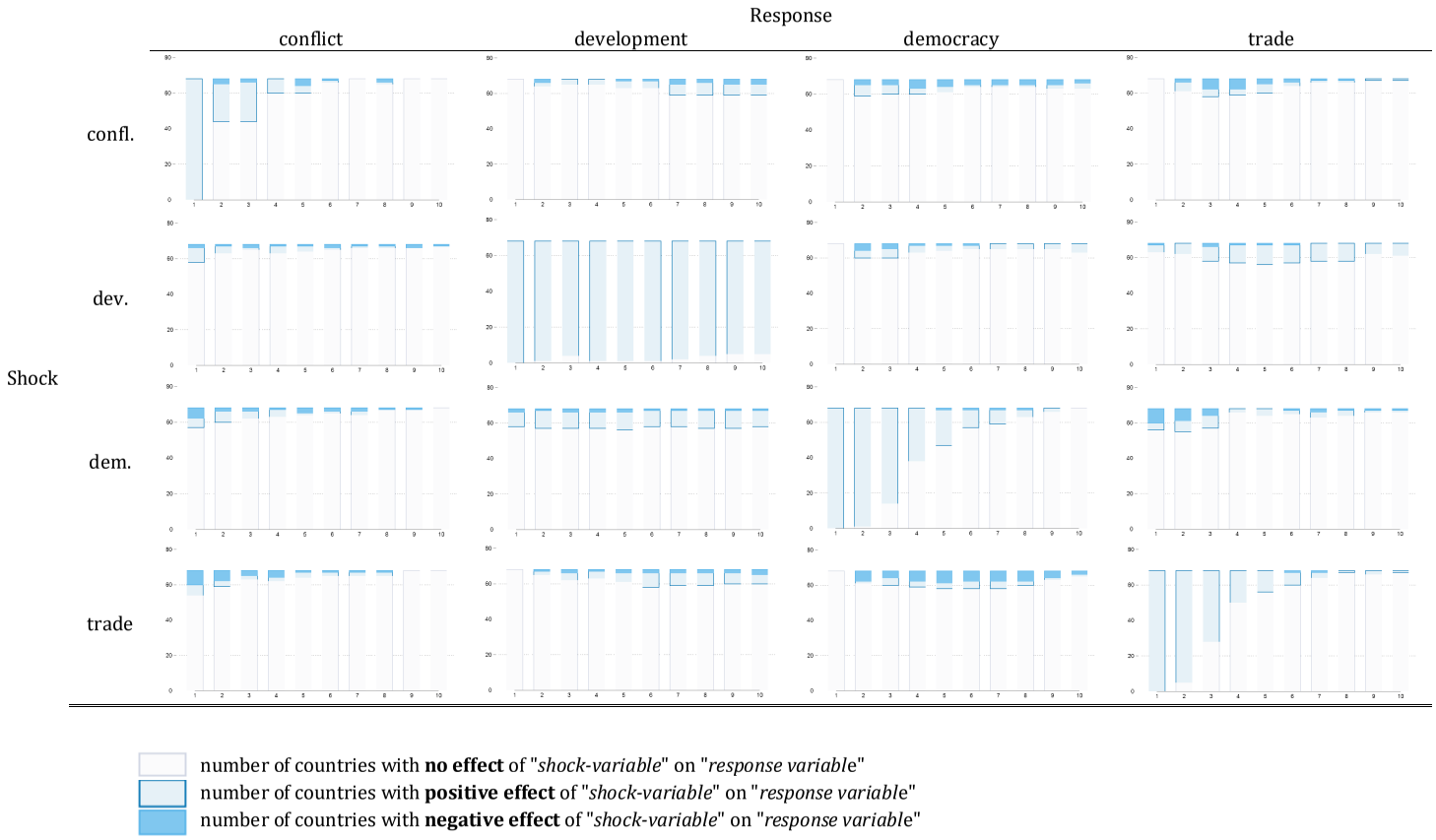


Figure D.27: Robustness Check - Model with alternative ordering (democracy, development, trade and conflict): Number of countries with non/significant responses to the given shocks

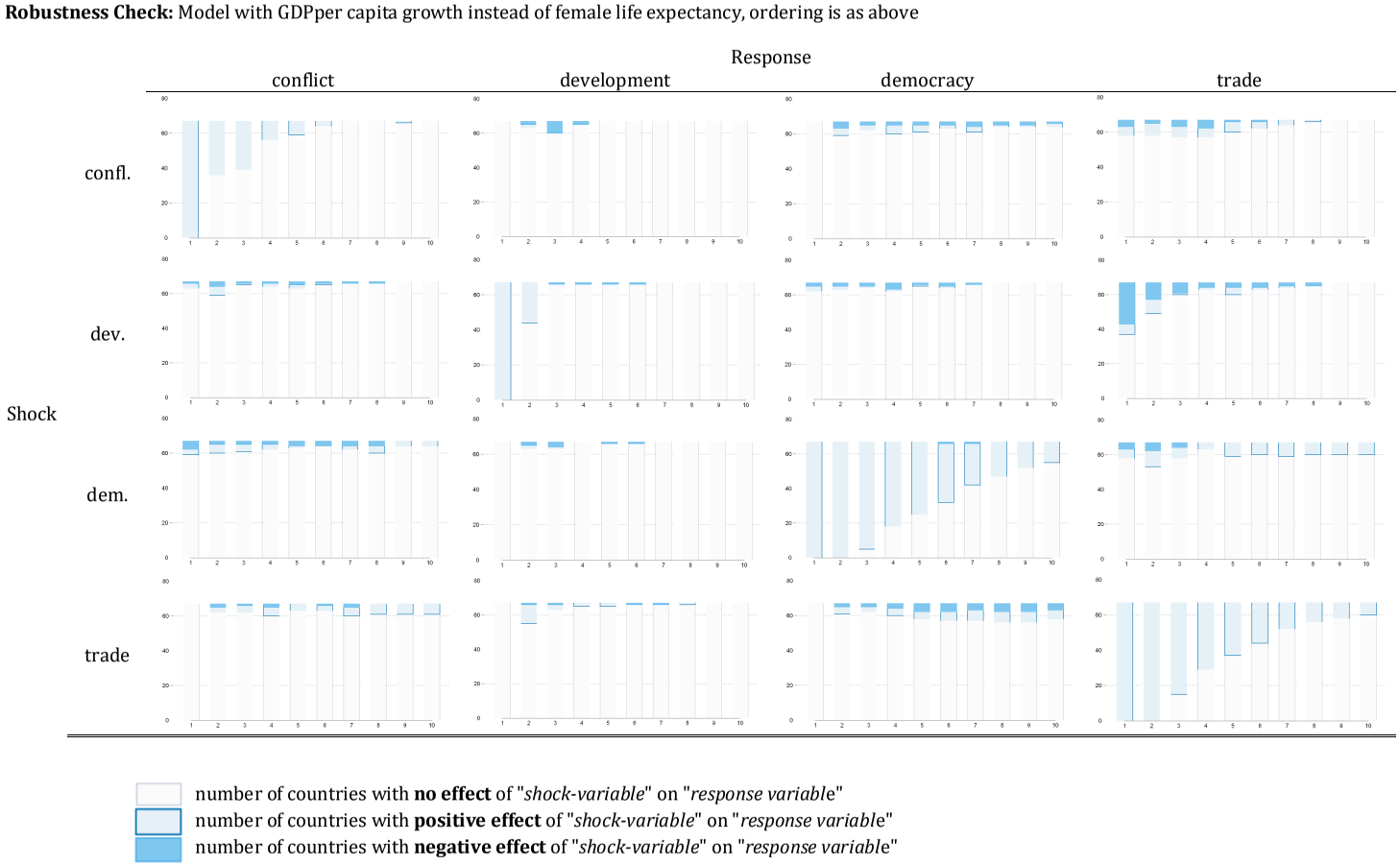


Figure D.28: Robustness Check - Model with GDP per capita growth: Number of countries with non/significant responses to the given shocks

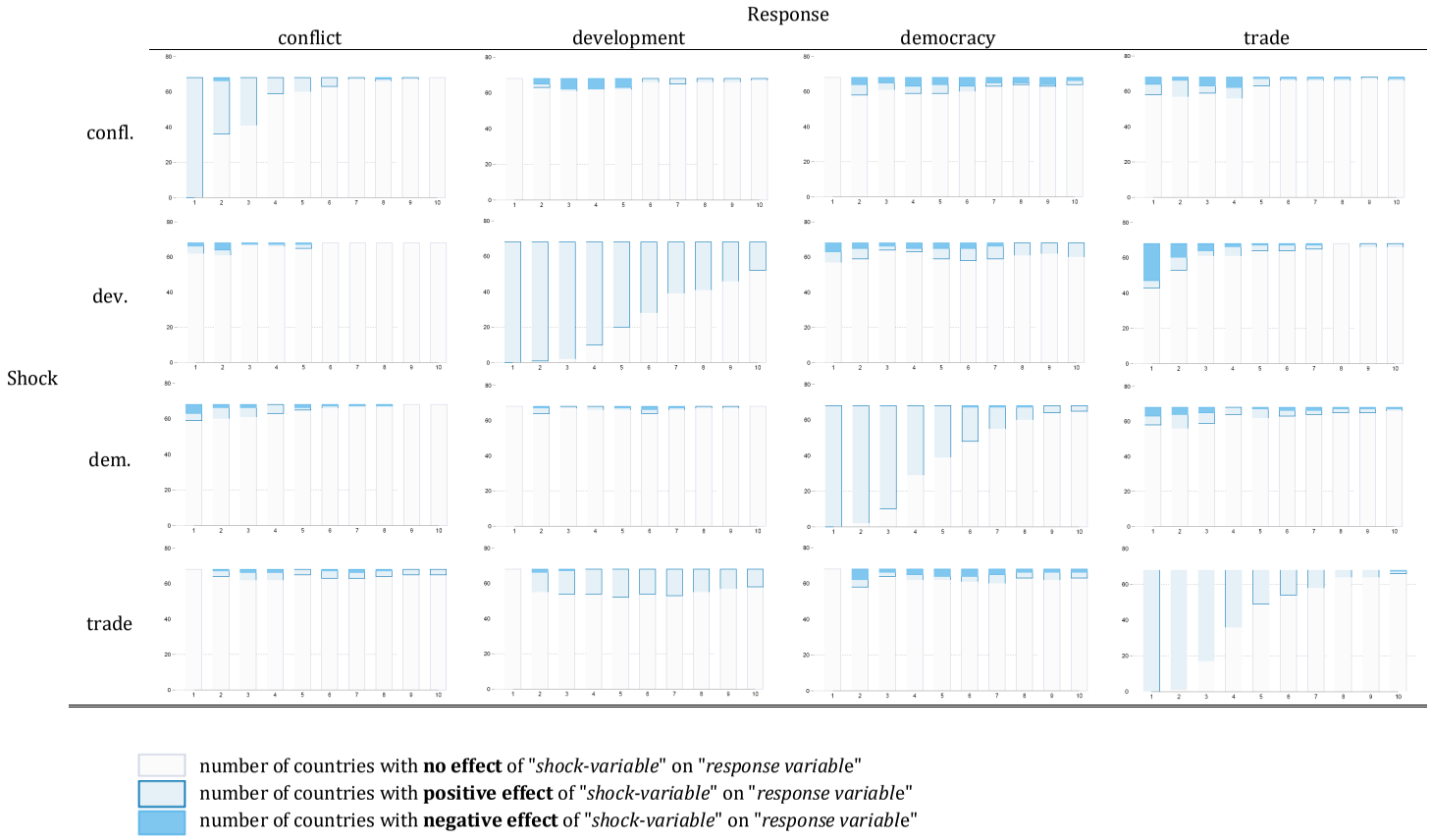


Figure D.29: Robustness Check - Model with logged GDP per capita: Number of countries with non/significant responses to the given shocks

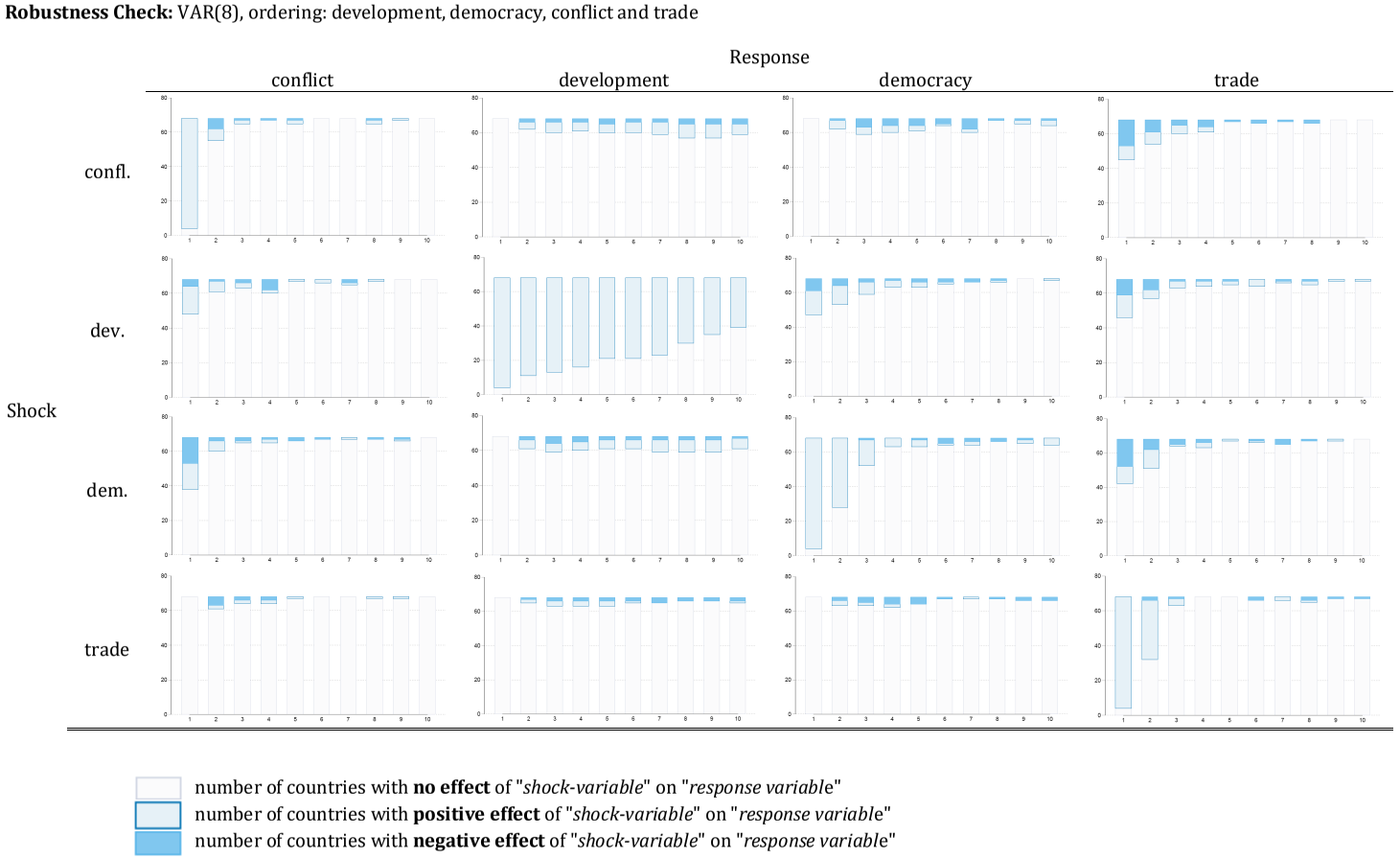


Figure D.30: Robustness Check - Model with lag length 8: Number of countries with non/significant responses to the given shocks

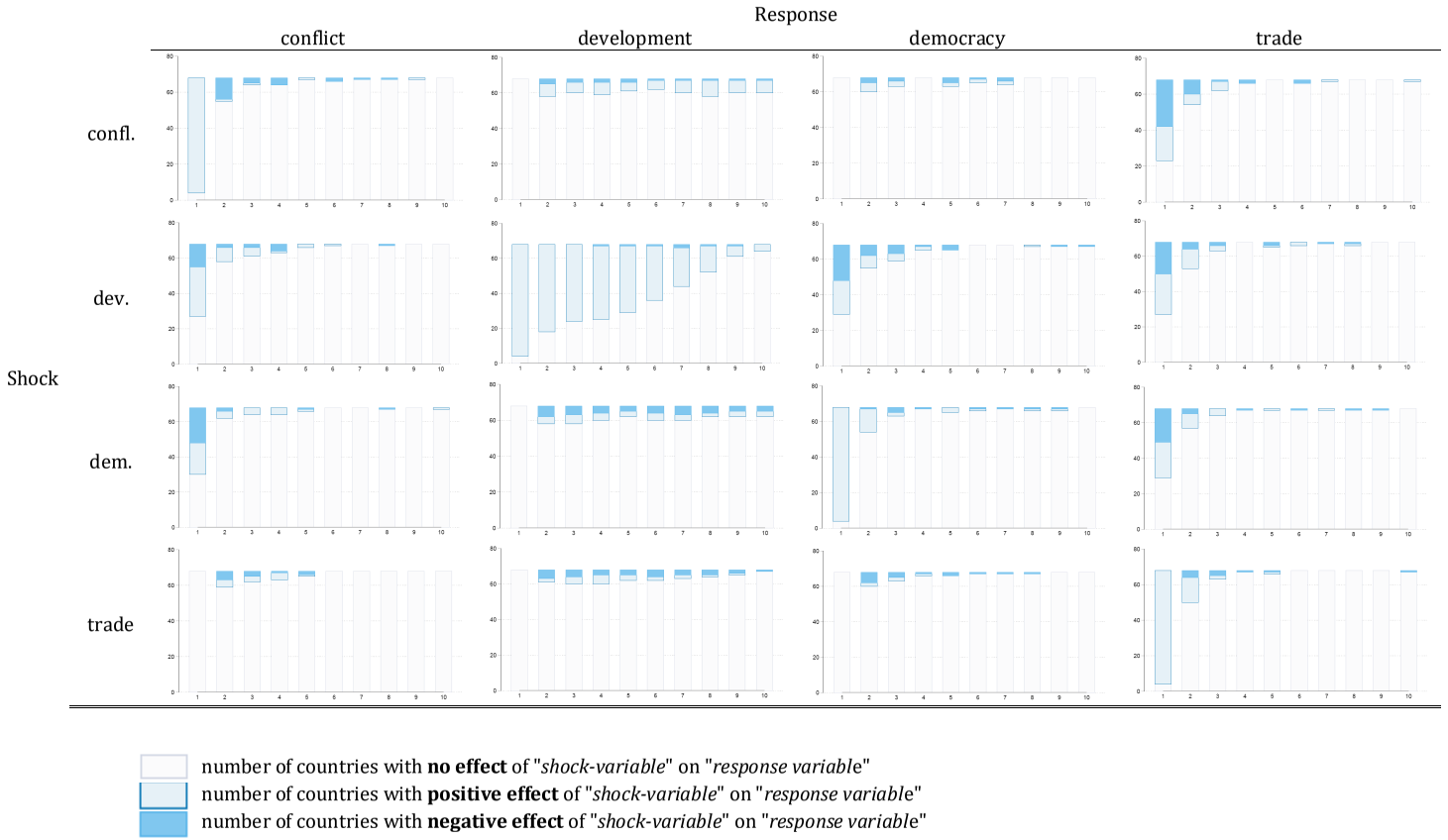


Figure D.31: Robustness Check - Model with lag length 10: Number of countries with non/significant responses to the given shocks

D.5 Examining residual autocorrelation

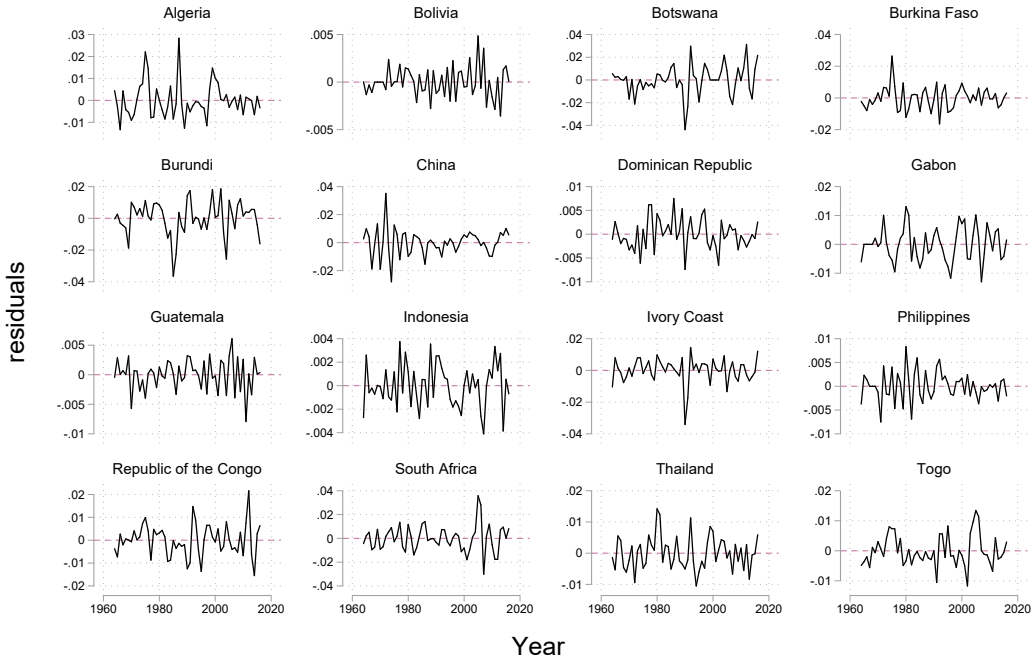


Figure D.32: VAR(4): residual plots for countries with LM-test p-values between .01-.05

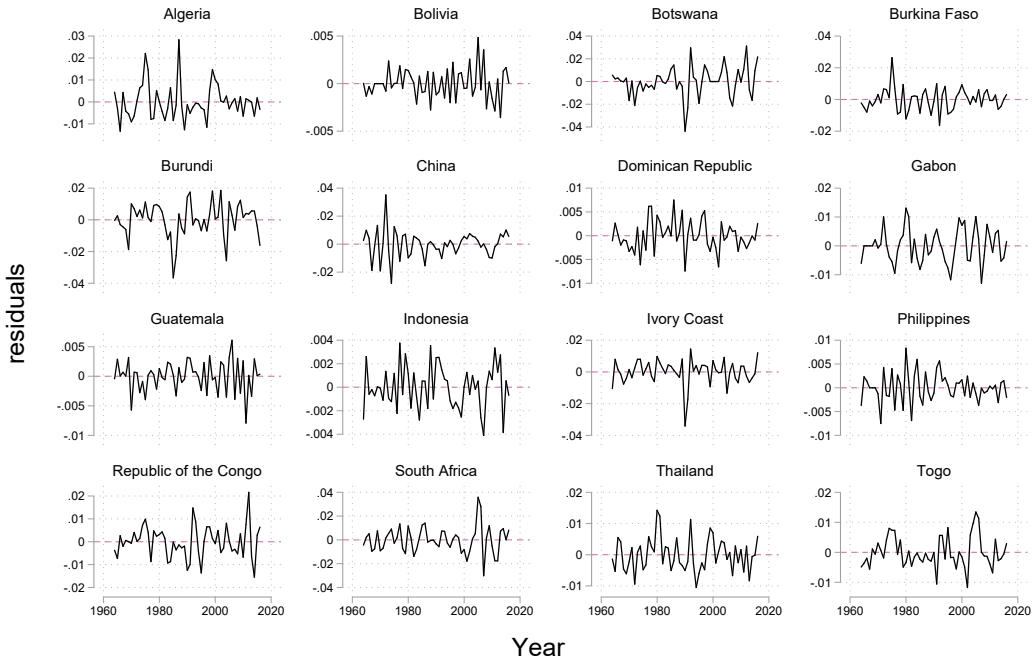


Figure D.33: VAR(4): residual plots for countries with LM-test p-values below .01

Bibliography

Acemoglu, Daron and James A Robinson (2001). "A theory of political transitions". In: *American Economic Review*, pp. 938–963.

— (2005). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Acemoglu, Daron et al. (2008). "Income and democracy". In: *American Economic Review* 98.3, pp. 808–42.

Acemoglu, Daron et al. (2014). *Democracy does cause growth*. Tech. rep. National Bureau of Economic Research.

— (2019). "Democracy does cause growth". In: *Journal of Political Economy* 127.1, pp. 47–100.

Adsera, Alicia and Carles Boix (2002). "Trade, democracy, and the size of the public sector: The political underpinnings of openness". In: *International Organization* 56.2, pp. 229–262.

Aghion, Philippe, Alberto F Alesina, and Francesco Trebbi (2007). "Democracy, technology, and growth". In: *National Bureau of Economic Research Cambridge, Mass., USA*.

Altman, David, Federico Rojas-de Galarreta, and Francisco Urdinez (2018). "An Interactive Model of the Democratic Peace: Revisiting the Theory with Elastic Measures". In:

Anderson, James E and Douglas Marcouiller (2002). "Insecurity and the pattern of trade: An empirical investigation". In: *Review of Economics and statistics* 84.2, pp. 342–352.

Anderson, James E and Eric Van Wincoop (2003). "Gravity with gravitas: a solution to the border puzzle". In: *American Economic Review* 93.1, pp. 170–192.

Anderton, Charles H and John R Carter (2009). *Principles of conflict economics: a primer for social scientists*. Cambridge University Press.

Barbieri, Katherine and Jack S Levy (1999). "Sleeping with the enemy: The impact of war on trade". In: *Journal of peace research* 36.4, pp. 463–479.

Barbieri, Katherine, Omar MG Keshk, and Brian M Pollins (2009). "Trading data evaluating our assumptions and coding rules". In: *Conflict Management and Peace Science* 26.5, pp. 471–491.

Barro, Robert J (1996). "Democracy and growth". In: *Journal of Economic Growth* 1.1, pp. 1–27.

Barro, Robert J. and Jong Wha Lee (2013). "A new data set of educational attainment in the world, 1950–2010". In: *Journal of development economics* 104, pp. 184–198.

Baum, Matthew A and David A Lake (2003). "The political economy of growth: democracy and human capital". In: *American Journal of Political Science* 47.2, pp. 333–347.

Bernhard, Michael, Ömer Faruk Örsün, and Reşat Bayer (2017). "Democratization in Conflict Research: How Conceptualization Affects Operationalization and Testing Outcomes". In: *International Interactions* 43.6, pp. 941–966.

Besley, Timothy and Masayuki Kudamatsu (2006). "Health and democracy". In: *American Economic Review* 96.2, pp. 313–318.

Besley, Timothy and Torsten Persson (2010). "State capacity, conflict, and development". In: *Econometrica* 78.1, pp. 1–34.

Bischof, Daniel et al. (2017). "New graphic schemes for Stata: plotplain and plottig". In: *Stata Journal* 17.3, pp. 748–759.

Blomberg, S Brock and Gregory D Hess (2006). "How much does violence tax trade?" In: *The Review of Economics and Statistics* 88.4, pp. 599–612.

Boehmer, Charles R, Bernadette ME Jungblut, and Richard J Stoll (2011). "Tradeoffs in trade data: Do our assumptions affect our results?" In: *Conflict Management and Peace Science* 28.2, pp. 145–167.

Boese, Vanessa A. (2015). "Viva la Revolución, or: Do Revolutions Lead to More Democracy?" In: *Peace Economics, Peace Science and Public Policy* 21, pp. 541 –551.

— (2019). "How (not) to measure democracy". In: *International Area Studies Review* 22, 95 – 127.

Boese, Vanessa A. and Katrin Kamin (2018a). "Democracy Datasets.xslx". In: *Available at: https://www.dropbox.com/s/3bm674tjk9iqha4/Democracy%20Datasets.xlsx?dl=0.*.

— (2018b). "Economic Datasets.xslx". In: *Available at: https://www.dropbox.com/s/qtaol2upv0uns30/Economic%20Datasets.xlsx.*

Boese, Vanessa A. et al. (2019). "Patterns of democracy over space and time". In: *Conference paper, 60th Annual Convention of the International Studies Association, 27-30 March 2019, Toronto, Ontario.*

Boese, Vanessa Alexandra and Katrin Kamin (2018c). "'Tis but thy name that is my enemy: on the construction of macro panel datasets in peace economics". In: *Economics of Peace and Security Journal* (forthcoming).

— (2019). "'Tis but thy name that is my enemy: on the construction of macro panel datasets in peace economics". In: *Economics of Peace and Security Journal* 14, pp. 5 –26.

Bogaards, Matthijs (2010). "Measures of democratization: From degree to type to war". In: *Political Research Quarterly* 63.2, pp. 475–488.

Boix, Carles, Michael Miller, and Sebastian Rosato (2013). "A complete data set of political regimes, 1800–2007". In: *Comparative Political Studies* 46.12, pp. 1523–1554.

Bollen, Kenneth A (1980). "Issues in the comparative measurement of political democracy". In: *American Sociological Review* 45.3, pp. 370–390.

Bollen, Kenneth A and Robert W Jackman (1985). "Political democracy and the size distribution of income". In: *American Sociological Review* 50.4, pp. 438–457.

— (1989). "Democracy, stability, and dichotomies". In: *American Sociological Review* 54.4, pp. 612–621.

Bollen, Kenneth A and Pamela Paxton (2000). "Subjective measures of liberal democracy". In: *Comparative Political Studies* 33.1, pp. 58–86.

Bollen, Kenneth (1993). "Liberal democracy: Validity and method factors in cross-national measures". In: *American Journal of Political Science* 37.4, pp. 1207–1230.

Boucekkine, Raouf, Paolo G Piacquadio, and Fabien Prieur (2016). "A Lipsetian Theory of Democratization: Development, Education, Inequality, and Resources". In:

Bush, Sarah Sunn (2017). "The politics of rating freedom: Ideological affinity, private authority, and the Freedom in the World ratings". In: *Perspectives on Politics* 15.3, pp. 711–731.

Casper, Gretchen and Claudiu Tufis (2003). "Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated data sets". In: *Political Analysis* 11.2, pp. 196–203.

Cervellati, Matteo et al. (2014). "Income and democracy: Comment". In: *American Economic Review* 104.2, pp. 707–19.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland (2010). "Democracy and dictatorship revisited". In: *Public choice* 143.1-2, pp. 67–101.

Chenoweth, Erica and Jay Ulfelder (2017). "Can structural conditions explain the onset of nonviolent uprisings?" In: *Journal of Conflict Resolution* 61.2, pp. 298–324.

Collier, Paul and Anke Hoeffler (2004). "Greed and grievance in civil war". In: *Oxford economic papers* 56.4, pp. 563–595.

Collier, Paul and Dominic Rohner (2008). "Democracy, development, and conflict". In: *Journal of the European Economic Association* 6.2-3, pp. 531–540.

Collier, Paul et al. (2003). *Breaking the conflict trap: Civil war and development policy*. World Bank Publications.

Coppedge, Michael, Angel Alvarez, and Claudia Maldonado (2008). "Two persistent dimensions of democracy: Contestation and inclusiveness". In: *The Journal of Politics* 70.3, pp. 632–647.

Coppedge, Michael et al. (2011). "Conceptualizing and measuring democracy: A new approach". In: *Perspectives on Politics* 9.2, pp. 247–267.

Coppedge, Michael et al. (2017a). "V-Dem Codebook V7". In: *V-Dem Reference Material (2017). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968274 (accessed on Dec. 21, 2017).*

Coppedge, Michael et al. (2017b). "V-Dem Comparisons and Contrasts with Other Measurement Projects". In: *V-Dem working paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2951014# (accessed on Dec. 21, 2017)* 2017.45.

Coppedge, Michael et al. (2017c). "V-Dem Country Coding Units V7". In: *V-Dem Reference Material (2017). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968285 (accessed on Dec. 21, 2017).*

Coppedge, Michael et al. (2017d). "V-Dem Dataset v7". In: *Available at: https://www.v-dem.net/en/data/data-version-7-1/ (accessed on Dec. 21, 2017).*

Coppedge, Michael et al. (2017e). "V-Dem Methodology v7". In: *V-Dem Reference Material (2017). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968284 (accessed on Dec. 21, 2017).*

Coppedge, Michael et al. (2017f). "V-Dem Organization and Management V7". In: *V-Dem Reference Material (2017). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968280 (accessed on Dec. 21, 2017).*

Coppedge, Michael et al. (2018a). "V-Dem Codebook V8". In:

Coppedge, Michael et al. (2018b). "V-Dem Country Coding Units v8". In:

Crescenzi, Mark JC and Kelly M Kadera (2016). "Built to Last: Understanding the Link between Democracy and Conflict in the International System". In: *International Studies Quarterly* 60.3, pp. 565–572.

Crick, Bernard (2002). *Democracy: a very short introduction*. Vol. 75. Oxford Paperbacks, Oxford University Press, Oxford.

Dahl, Robert Alan (1971). *Polyarchy: Participation and opposition*. Yale University Press, New Haven and London.

Diamond, Larry (1999). *Developing democracy: Toward consolidation*. Johns Hopkins University Press, Baltimore Maryland, USA.

Doorenspleet, Renske (2000). "Reassessing the three waves of democratization". In: *World Politics* 52.03, pp. 384–406.

Doucouliagos, Hristos and Mehmet Ali Ulubaşoğlu (2008). "Democracy and economic growth: a meta-analysis". In: *American Journal of Political Science* 52.1, pp. 61–83.

Dunne, J Paul and Nan Tian (2015). "Military expenditure, economic growth and heterogeneity". In: *Defence and Peace Economics* 26.1, pp. 15–31.

Eckstein, Harry (1973). "Authority patterns: A structural basis for political inquiry". In: *American Political Science Review* 67.4, pp. 1142–1161.

Egger, Hartmut and Udo Kreickemeier (2012). "Fairness, trade, and inequality". In: *Journal of International Economics* 86.2, pp. 184–196.

Elff, Martin and Sebastian Ziaja (2018). "Method factors in democracy indicators". In: *Politics and Governance* 6.1, pp. 92–104.

Elkins, Zachary (2000). "Gradations of democracy? Empirical tests of alternative conceptualizations". In: *American Journal of Political Science* 44.2, pp. 293–300.

Fearon, James D and David D Laitin (2003). "Ethnicity, insurgency, and civil war". In: *American political science review* 97.1, pp. 75–90.

Franco, Álvaro, Carlos Álvarez-Dardet, and Maria Teresa Ruiz (2004). "Effect of democracy on health: ecological study". In: *Bmj* 329.7480, pp. 1421–1423.

Frankel, Jeffrey A and David H Romer (1999). "Does trade cause growth?" In: *American economic review* 89.3, pp. 379–399.

Freedom House (2017a). "Freedom in the World 2016: Anxious Dictators, Wavering Democracies: Global Freedom under Pressure". In: *Available at: https://freedomhouse.org/sites/default/files/FH_FITW_Report_2016.pdf (accessed on Dec. 1, 2017).*

— (2017b). "Methodology Website". In: *Available at: https://freedomhouse.org/report/freedom-world-2016/methodology (accessed on Dec. 1, 2017).*

Gartzke, Erik and Alex Weisiger (2014). "Under construction: Development, democracy, and difference as determinants of systemic liberal peace". In: *International Studies Quarterly* 58.1, pp. 130–145.

Gates, Scott, Torbjørn L Knutsen, and Jonathon W Moses (1996). "Democracy and peace: A more skeptical view". In: *Journal of Peace Research* 33.1, pp. 1–10.

Gates, Scott et al. (2006). "Institutional inconsistency and political instability: Polity duration, 1800–2000". In: *American Journal of Political Science* 50.4, pp. 893–908.

Giavazzi, Francesco and Guido Tabellini (2005). "Economic and political liberalizations". In: *Journal of Monetary Economics* 52.7, pp. 1297–1330.

Gleditsch, Kristian S and Michael D Ward (1997). "Double take: A reexamination of democracy and autocracy in modern polities". In: *Journal of Conflict Resolution* 41.3, pp. 361–383.

— (1999). "A revised list of independent states since the Congress of Vienna". In: *International Interactions* 25.4, pp. 393–413.

Gleditsch, Nils Petter et al. (2002). "Armed conflict 1946-2001: A new dataset". In: *Journal of peace research* 39.5, pp. 615–637.

Glick, Reuven and Alan M Taylor (2010). "Collateral damage: Trade disruption and the economic impact of war". In: *The Review of Economics and Statistics* 92.1, pp. 102–127.

Goertz, Gary (2006). *Social science concepts: A user's guide*. Princeton University Press, Princeton, New Jersey, USA.

Goldstone, Jack A et al. (2010). "A global model for forecasting political instability". In: *American Journal of Political Science* 54.1, pp. 190–208.

Gurr, Ted Robert (1974). "Persistence and change in political systems, 1800–1971". In: *American Political Science Review* 68.4, pp. 1482–1504.

Gurr, Ted Robert, Monty G. Marshall, and Keith Jaggers (2013). "Polity IV Project: Dataset Users' Manual - Political regime characteristics and transitions, 1800-2012". In: *Available at http://www.systemicpeace.org/inscrdata.html. Accessed on July 31, 2015.*

Hegre, Håvard (2001). "Toward a democratic civil peace? Democracy, political change, and civil war, 1816–1992". In: *American political science review* 95.1, pp. 33–48.

— (2014). "Democracy and armed conflict". In: *Journal of Peace Research* 51.2, pp. 159–172.

Hegre, Håvard, John R Oneal, and Bruce Russett (2010). "Trade does promote peace: New simultaneous estimates of the reciprocal effects of trade and conflict". In: *Journal of Peace Research* 47.6, pp. 763–774.

Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding (2010). "Inequality and unemployment in a global economy". In: *Econometrica* 78.4, pp. 1239–1283.

Hensel, Paul (2001). "ICOW Historical State Names Data Set, Version 1.2". In: *Issue Correlates of War (ICOW) Project. Available at: http://www.paulhensel.org/icownames.html.*

Heston, Alan, Robert Summers, and Bettina Aten (2011). "Penn World Table Version 7.0". In: *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.*

Högström, John (2013). "Does the choice of democracy measure matter? Comparisons between the two leading democracy indices, Freedom House and Polity IV". In: *Government and Opposition* 48.2, pp. 201–221.

Høyland, Bjørn, Karl Moene, and Fredrik Willumsen (2012). "The tyranny of international index rankings". In: *Journal of Development economics* 97.1, pp. 1–14.

Huntington, Samuel P. (1993). *The third wave: Democratization in the late twentieth century*. Vol. 4. University of Oklahoma Press, Norman, USA.

Inkeles, Alex (1991). *On measuring democracy: Its consequences and concomitants*. Transaction Publishers, New Brunswick, New Jersey, USA.

Jaggers, Keith and Ted Robert Gurr (1995). "Tracking democracy's third wave with the Polity III data". In: *Journal of peace research* 32.4, pp. 469–482.

Justino, Patricia, Tilman Brück, and Philip Verwimp (2013). *A micro-level perspective on the dynamics of conflict, violence, and development*. Oxford University Press.

Kamin, Katrin (n.d.). "Is conflict really a trade barrier? Assessing the impact of different conflict types on trade with a ppml panel approach". In:

Keesing, Donald B (1967). "The impact of research and development on United States trade". In: *Journal of Political Economy* 75.1, pp. 38–48.

Keshk, Omar MG, Rafael Reuveny, and Brian M Pollins (2010). "Trade and conflict: Proximity, country size, and measures". In: *Conflict Management and Peace Science* 27.1, pp. 3–27.

Knutsen, Carl Henrik (2010). "Measuring effective democracy". In: *International Political Science Review* 31.2, pp. 109–128.

— (2011). *The economic effects of democracy and dictatorship*. PhD Thesis, University of Oslo, Norway.

— (2013). "Democracy, state capacity, and economic growth". In: *World Development* 43, pp. 1–18.

— (2015). "Reinvestigating the reciprocal relationship between democracy and income inequality". In: *Review of Economics and Institutions* 6.2, pp. 1–37.

Lake, David A and Matthew A Baum (2001). "The invisible hand of democracy: political control and the provision of public services". In: *Comparative Political Studies* 34.6, pp. 587–621.

Li, Quan and Rafael Reuveny (2003). "Economic globalization and democracy: An empirical analysis". In: *British Journal of Political Science* 33.1, pp. 29–54.

Lindberg, Staffan I et al. (2014). "V-Dem: A new way to measure democracy". In: *Journal of Democracy* 25.3, pp. 159–169.

Lipset, Seymour Martin (1959). "Some social requisites of democracy: Economic development and political legitimacy". In: *American Political Science Review* 53.1, pp. 69–105.

Liu, Xuepeng and Emanuel Ornelas (2014). "Free trade agreements and the consolidation of democracy". In: *American Economic Journal: Macroeconomics* 6.2, pp. 29–70.

Long, Andrew G (2008). "Bilateral trade in the shadow of armed conflict". In: *International Studies Quarterly* 52.1, pp. 81–101.

Mac Ginty, Roger and Andrew Williams (2016). *Conflict and development*. Routledge.

Madsen, Jakob B, Paul A Raschky, and Ahmed Skali (2015). "Does democracy drive income in the world, 1500–2000?" In: *European Economic Review* 78, pp. 175–195.

Mainwaring, Scott, Daniel Brinks, and Aníbal Pérez-Liñán (2007). "Classifying political regimes in Latin America, 1945-2004". In: *In "Regimes and democracy in Latin America: Theories and methods". Edited by Gerardo L. Munck. Oxford University Press, Oxford, UK*, pp. 123–160.

Mansfield, Edward D and Jack Snyder (2002). "Democratic transitions, institutional strength, and war". In: *International Organization* 56.2, pp. 297–337.

Marquardt, Kyle L and Daniel Pemstein (2018). "IRT models for expert-coded panel data". In: *Political Analysis* 26.4, pp. 431–456.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers (2013). "Polity IV Project: Political regime characteristics and transitions, 1800-2012". In: *Available at* `http://www.systemicpeace.org/inscrdata.html`. *Accessed on June 30th, 2014*.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers (2017a). "Polity IV Project: Dataset Users' Manual - Political regime characteristics and transitions, 1800-2016". In: *Available at http://www.systemicpeace.org/inscrdata.html. Accessed on July 6th, 2017*.

— (2017b). "Polity IV Project: Political regime characteristics and transitions, 1800-2016". In: *Available at http://www.systemicpeace.org/inscrdata.html (accessed on July 6, 2017)*.

Marshall, Monty (2014). *Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2013, Center for Systemic Peace*.

Martin, Philippe, Thierry Mayer, and Mathias Thoenig (2008). "Make trade not war?" In: *The Review of Economic Studies* 75.3, pp. 865–900.

Mitchell, Sara McLaughlin, Scott Gates, and Håvard Hegre (1999). "Evolution in democracy-war dynamics". In: *Journal of Conflict Resolution* 43.6, pp. 771–792.

Munck, Gerardo L (2009). *Measuring democracy: A bridge between scholarship and politics*. Johns Hopkins University Press, Baltimore, Maryland.

Munck, Gerardo L and Jay Verkuilen (2002). "Conceptualizing and measuring democracy: Evaluating alternative indices". In: *Comparative political studies* 35.1, pp. 5–34.

Murtin, Fabrice and Romain Wacziarg (2014). "The democratic transition". In: *Journal of Economic Growth* 19.2, pp. 141–181.

Narayan, Paresh Kumar, Seema Narayan, and Russell Smyth (2011). "Does democracy facilitate economic growth or does economic growth facilitate democracy? An empirical study of Sub-Saharan Africa". In: *Economic Modelling* 28.3, pp. 900–910.

Oneal, John R et al. (1996). "The liberal peace: Interdependence, democracy, and international conflict, 1950-85". In: *Journal of Peace Research* 33.1, pp. 11–28.

Palmer, Glenn et al. (2015). "The MID4 dataset, 2002–2010: Procedures, coding rules and description". In: *Conflict Management and Peace Science* 32.2, pp. 222–242.

Papaioannou, Elias and Gregorios Siourounis (2008). "Democratisation and growth". In: *The Economic Journal* 118.532, pp. 1520–1551.

Pemstein, Daniel, Stephen A Meserve, and James Melton (2010). "Democratic compromise: A latent variable analysis of ten measures of regime type". In: *Political Analysis* 18.4, pp. 426–449.

Persson, Torsten and Guido Tabellini (2009). "Democratic capital: The nexus of political and economic change". In: *American Economic Journal: Macroeconomics* 1.2, pp. 88–126.

Pettersson, Therése and Kristine Eck (2018). "Organized violence, 1989–2017". In: *Journal of Peace Research* 55.4, pp. 535–547.

Pozuelo, Julia, Amy Slipowitz, and Guillermo Vuletin (2016). "Democracy does not cause growth: The importance of endogeneity arguments". In:

Przeworski, Adam (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America*. Cambridge University Press.

Przeworski, Adam et al. (2000). *Democracy and development: political institutions and well-being in the world, 1950-1990*. Vol. 3. Cambridge University Press, Cambridge, UK.

Raciborski, Rafal (2008). "kountry: A Stata utility for merging cross-country data from multiple sources". In: *The Stata Journal* 8.3, pp. 390–400.

Reuveny, Rafael and Quan Li (2003). "Economic openness, democracy, and income inequality: an empirical analysis". In: *Comparative Political Studies* 36.5, pp. 575–601.

Rigobon, Roberto and Dani Rodrik (2005). "Rule of law, democracy, openness, and income: Estimating the interrelationships1". In: *Economics of Transition* 13.3, pp. 533–564.

Rodriguez, Francisco and Dani Rodrik (2000). "Trade policy and economic growth: a skeptic's guide to the cross-national evidence". In: *NBER macroeconomics annual* 15, pp. 261–325.

Rodrik, Dani and Romain Wacziarg (2005). "Do democratic transitions produce bad economic outcomes?" In: *American Economic Review* 95.2, pp. 50–55.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi (2004). "Institutions rule: the primacy of institutions over geography and integration in economic development". In: *Journal of Economic Growth* 9.2, pp. 131–165.

Russett, Bruce M, J David Singer, and Melvin Small (1968). "National political units in the twentieth century: A standardized list". In: *American Political Science Review* 62.3, pp. 932–951.

Russett, Bruce and John Oneal (2001). "Triangulating peace". In: *Democracy, Interdependence, and International Organizations, New York*.

Santos Silva, JMC and Silvana Tenreyro (2006). "The log of gravity". In: *The Review of Economics and Statistics* 88.4, pp. 641–658.

Schmitter, Philippe C and Terry Lynn Karl (1991). "What democracy is... and is not". In: *Journal of democracy* 2.3, pp. 75–88.

Schumpeter, Joseph A (1976 [1942]). *Capitalism, socialism and democracy*. Routledge: London and New York.

Shakespeare, William (2003). *romeo and juliet*. Cambridge University Press.

Strand, Havard et al. (2012). *Why Waves? Global Patterns of Democratization, 1816-2008*. Tech. rep. Mimeo, PRIO.

Strezhnev, Anton and Erik Voeten (2013). "United Nations general assembly voting data, Data and Codebook, Version 7". In:

Subramanian, Arvind and Shanker Satyanath (2004). *What determines long-run macroeconomic stability? Democratic institutions*. 4-215. International Monetary Fund.

Tavares, Jose and Romain Wacziarg (2001). "How democracy affects growth". In: *European Economic Review* 45.8, pp. 1341–1378.

The World Bank (2014). *The World Development Indicators, Database*. Washington, DC, USA. Available at `http://data.worldbank.org/data-catalog/world-development-indicators`. Assessed on March 26, 2015.

— (2017b). *World Development Indicators, Database*. Washington DC. Available at: `https://data.worldbank.org/products/wdi`. Assessed on March 26, 2015.

— (2017a). *World Development Indicators*. Annual Report, Washington DC. License: Creative Commons Attribution CC BY 3.0 IGO. Available at: `https://openknowledge.worldbank.org/handle/10986/26447`.

Themnér, Lotta and Peter Wallensteen (2014). "Armed Conflict, 1946-2013, Version 4-2014". In: *Journal of Peace Research* 51.4, pp. 541–554.

Themnér, Lotta (2018). "UCDP/PRIO Armed Conflict Dataset Codebook, Version 18.1". In: *Uppsala Conflict Data Program and International Peace Research Institute, Oslo. Available at: https://ucdp.uu.se/downloads/ucdpprio/ucdp-prio-acd-181.pdf*.

Treier, Shawn and Simon Jackman (2008). "Democracy as a latent variable". In: *American Journal of Political Science* 52.1, pp. 201–217.

"United Nations Comtrade Database" (2018). In: *DESA/UNSD. Available at: https://comtrade.un.org/*.

Vanhanen, Tatu (2000). "A new dataset for measuring democracy, 1810-1998". In: *Journal of Peace Research* 37.2, pp. 251–265.

Vreeland, James Raymond (2008). "The effect of political regime on civil war: Unpacking anocracy". In: *Journal of Conflict Resolution* 52.3, pp. 401–425.

Yanikkaya, Halit (2003). "Trade openness and economic growth: a cross-country empirical investigation". In: *Journal of Development economics* 72.1, pp. 57–89.

Yu, Miaojie (2010). "Trade, democracy, and the gravity equation". In: *Journal of Development Economics* 91.2, pp. 289–300.

d'Agostino, Giorgio, John Paul Dunne, and Luca Pieroni (2018). "Military expenditure, endogeneity and economic growth". In: *Defence and Peace Economics*, pp. 1–16.

Statement

I testify through my signature that all information that I have provided about resources used in the writing of my doctoral thesis, about the resources and support provided to me as well as in earlier assessments of my doctoral thesis correspond in every aspect to the truth.

Berlin, 15th April 2019

.....

(Signature: Vanessa Alexandra Boese)